

Time to withdraw an undesirable privilege?

The practice by which some researchers restrict access to published data for a year has hitherto been accepted. *Nature* and *Science* are collaborating to investigate whether that acceptance should cease.

A principled editorial policy, implemented with rigour, can lose a journal some excellent papers to other journals that are more lax. What is more, journals are in a poor position to wag the dog of community behaviour. Nevertheless, there are occasions when they can have a significant influence, especially if competitiveness can be sidestepped. One such occasion may be approaching in respect of the vexed issue of the availability, on publication, of crystallographic coordinates of biomolecular structures.

Most researchers believe that, once a piece of science has been published, the methods and materials on which it is based as well as the results themselves should be publicly available, immediately. Many structural biologists disagree. For them, it is a common practice, on publication, to deposit crystallographic atomic coordinates in a repository such as the protein database at Brookhaven National Laboratory, but to keep them inaccessible for a year.

In this context, *Nature's* policy is to publish when (and only when) a database accession number is supplied, but to accept the principle of the one-year embargo. But that acceptance is increasingly reluctant. As the use of structural data becomes ever greater across biology, that practice is ever more frustrating to the progress of science, not to say offensive to the broader community.

One obstacle to change is the claim by some structural biologists (see *Nature Structural Biology* 5, 83; 1998) that researchers are owed the breathing space of a one-year hold to exploit results that have been obtained by dint of investment of many months of work, with the substantial risk that the analysis will prove unsuccessful. Researchers in other fields may well view such a claim as special

pleading. Another obstacle is the knowledge that such data may not be patentable (although that possibility is being investigated by lawyers), and also that industry, with major quantities of information up its sleeve that non-industrial researchers can make good use of, will refuse to collaborate if immediate openness is required.

The extent to which all of these considerations should determine policy is highly debatable. Nevertheless, they clearly win the argument within the International Union of Crystallographers, whose continuing support for the one-year hold has considerable influence within the structural biological community.

Nature is a service provider. If the journal singlehandedly insists on immediate public access, and enforces that as rigorously as it does, for example, for geneticists, it could lose some excellent papers — or, to put it less selfishly, could cease to be considered as an appropriate service provider. But whose interests are we serving? It seems that increasingly, in accepting the one-year hold, we are serving industry at the expense of science.

We believe that the best way forward is to consult but also to minimize the use of such policy as a competitive weapon between journals. Accordingly, *Nature* and *Science* have agreed to collaborate in consulting the community and, potentially, reaching a simultaneous and mutually consistent decision on any change of policy. (Other journals are also being consulted.) Both journals are encouraging readers to express their opinions via a survey being conducted via the website of *Nature Structural Biology* (<http://structbio.nature.com>). Either by that means or by correspondence (corres@nature.com), we would welcome readers' opinions. □

Stockpiler's suppression

Is the US Department of Energy abusing its monopoly of information on nuclear weapons to stifle argument?

One reason why the United States elected to place its nuclear weapons technology under the control of a civilian agency was that issues surrounding the weapons were seen to be too important to be left to a military clique. The resulting system does not allow information on nuclear weapons to pour out into the public domain. Important technical information is kept secret — but it is nonetheless made available to a fairly wide circle of security-cleared experts, who are able to use it to inform a relatively open public debate on weapons policy.

There are disquieting signs that the Department of Energy (DOE) is failing to adhere to this mode of operation, as it singlemindedly pursues its intended policy of science-based stockpile stewardship — a programme developed to maintain the viability of existing US nuclear weapons after the proposed implementation of the Comprehensive Test Ban Treaty.

Ray Kidder, a weapons designer with a long and influential track record in both theoretical physics and nuclear weapons policy, has been denied the opportunity to help the Arms Control and Disarma-

ment Agency (ACDA) to assess maintenance options for nuclear weapons (see page 622). The denial came after the energy department declined to give him the necessary access to classified material. ACDA is not strong at the moment — it is about to be absorbed by the State Department — and it has not been able to stand up to DOE. The National Security Council may also have intervened to secure Kidder's exclusion.

But concern about the closed nature of the debate on stockpile stewardship does not stop with the treatment of Ray Kidder, who made his critical views on the issue known in *Nature* last year (386, 645; 1997). The formulation of the entire policy during the Clinton administration has been overly constrained by political concerns. No-one outside DOE has been allowed to compare stewardship — a crafty if expensive compromise solution — with alternative approaches, such as the remanufacturing option advocated by Kidder. Congress should insist on such a comparison, as it considers the administration's extravagant request of \$4.5 billion for stockpile stewardship in 1999. □



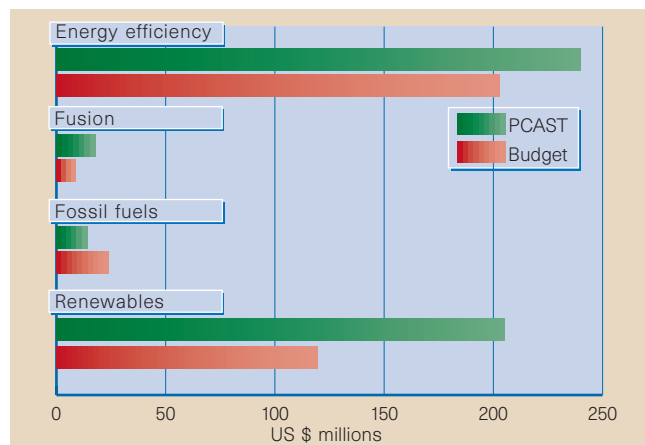
\$6bn package to cut US carbon emissions comes under attack

[WASHINGTON] Energy policy experts are sharply divided about the adequacy of a proposed research and tax incentive package, worth \$6.3 billion over five years, that the Clinton administration says is needed for the United States to comply with the Kyoto agreement on climate change.

The package — part of the budget proposals announced last week — has been welcomed by supporters of the Kyoto deal. It is the administration's first action to comply with the agreement, and may be its only action in advance of the agreement's ratification by the US Senate.

But economists, including those at the Energy Information Administration (EIA), an independent branch of the government charged with assessing such matters, doubt whether the package will have much impact on total US emissions of carbon dioxide. According to the EIA, these will exceed their 1990 levels by 34 per cent by 2010; the Kyoto accord requires emissions reductions of 7 per cent by then.

At a hearing today (12 February) of the Science Committee of the House of Representatives, chaired by James Sensenbrenner



Two views: contrasting increases proposed between 1997 and 1999 for applied energy research by the President's Council of Advisors on Science and Technology, and in the Clinton administration's 1999 budget request.

(Republican, Wisconsin), Republicans may launch an attack on the package that could further diminish its impact.

The package consists of \$3.7 billion of tax incentives for businesses and consumers, and \$2.7 billion of spending on research and development of technologies to cut carbon emissions, over the period 1999–2003.

The tax incentives include credits of \$3,000 for individuals buying big cars that use less fuel — there will be no incentive to

buy smaller vehicles, says the administration — and a batch of similar credits for everything from solar roof panels to full-blown combined heat and power systems.

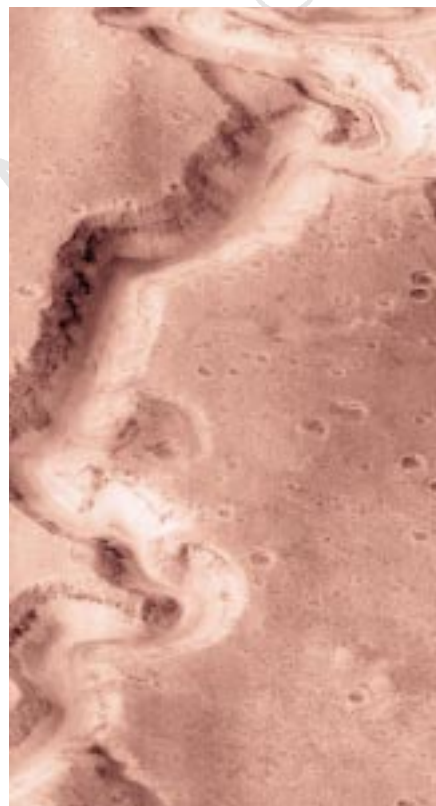
The research and development proposals are broadly based on recommendations made last November by a panel of the President's Council of Advisors on Science and Technology, chaired by John Holdren of the John F. Kennedy School of Government at Harvard University. Its report called for extra investment in energy supply research at the Department of Energy (DOE) of \$370 million this year, rising to \$1 billion by 2003.

It is difficult to compare what Holdren asked for with what the budget delivers. The report based its recommended increases on a level of spending for 1998 that was subsequently cut by the Congress. For example, Holdren suggested that research in solar and renewable energy be increased next year by \$130 million, to \$475 million. The Clinton budget does ask for a substantial increase, of \$90 million. But, because the baseline has been cut back, the Clinton proposal increases spending to just \$390 million. In other words, the budget proposal goes halfway to meeting the Holdren panel's expectations.

Overall, the Holdren report envisaged extra spending of \$3.8 billion at DOE over the five years from 1999 to 2003. The Clinton budget proposes extra expenditure of \$1.9 billion at the department over the period — again, half what the report asked for. The rest of the \$2.7 billion research and development package proposed by Clinton is spent in other agencies, chiefly the Environmental Protection Agency (EPA).

A spokesman for the EPA was unable to provide a detailed breakdown of this expenditure. But briefings to Congress indicate that the extra funds will go to technology partnership programmes between the

Clues to the origins of canyons on Mars



[WASHINGTON] Mystery veils the origin of the 2.5-km-wide canyon of Nanedi Vallis, one of the Martian valley systems — here pictured by the Mars Orbiter Camera. The picture was taken last month during the 87th orbit of the Mars Global Surveyor spacecraft.

Rocky outcrops occur along the upper walls, and weathered debris is found on the lower slopes and along the floor of the canyon, which cuts through cratered plains in the Xanthe Terra region of Mars.

Some features, such as terraces within the canyon and the 200-metre wide channel — both evident near the top of the picture — suggest it arose from continual fluid flow. Others, such as the lack of a contributing pattern of smaller channels on the surface surrounding the canyon, the shape of tributaries and the size and tightness of the apparent meanders, indicate that it was formed by collapse.

It is now thought likely that both continual flow and collapse were responsible for the canyon's appearance. Further observations are now planned to help to distinguish the relative effect of these and other potential formation and modification processes.

agency and industry. Such programmes have, in the past, been strongly resisted by Republicans in the Congress.

If the package is implemented, spending on energy supply research and development will still be at historically low levels. This investment peaked at almost \$10 billion, in today's money, in 1978.

The Holdren report sought to lift it from a 1997 trough of just \$1.3 billion. Mike Doyle (Democrat, Pennsylvania), a member of the science committee and supporter of coal research, says: "\$2.7 billion [over five years] is a pittance. We should be spending maybe ten times as much if we're serious about this."

Testifying before the science committee last week, Jay Hakes, administrator of the EIA, said that the long life of cars, buildings and factories ensured that there were "no quick fixes" to reducing emissions.

EIA already assumes improvements in energy efficiency in its projections, he said, adding that changes in those assumptions resulting from new technology "don't come close" to stabilizing US carbon emissions. "It's unlikely that any reduction can be achieved without a price mechanism," he said.

Government officials argue that the new energy research effort is more focused than before. They say the old programmes involved large and unproductive 'technology demonstrators'. The new programmes — such as a Nuclear Energy Research Initiative, recommended by Holdren — allow universities and government laboratories to compete for funds under carefully supervised peer review.

But environmentalists have nevertheless declared war on the new nuclear initiative. And according to congressional staff — even liberal Democrats — the DOE programme contains items that look more like industry support than peer-reviewed research.

As the proposal passes through Congress, its fate will depend on the wider budget picture and on Republican sentiment towards Kyoto. At last week's science committee hearing, some Republicans expressed support for energy research. Roscoe Bartlett (Republican, Maryland) said he did not believe in manmade global warming, but still wanted research on alternatives because they are needed to replace fossil fuels.

Alden Meyer, a lobbyist for the Union of Concerned Scientists, which supports strong action to reduce emissions, hopes and believes that the administration will fight hard to keep the package in the budget. "The question is whether the Republicans will oppose it, as a proxy vote against the Kyoto deal, or support it as something in line with their view that a carrot, and not a stick, should be used" to cut emissions, said Meyer. "I think they are split, currently." But, even if the package prevails, there is little evidence that it will help the United States to comply with its Kyoto target.

Colin Macilwain

Physicists oppose plan to slim down fusion reactor...

[MUNICH] European fusion scientists are determined to stick to the original ambitions of the 'next-step' fusion reactor, ITER, despite growing concern that the United States may back away, and waning enthusiasm from some European countries.

ITER (International Thermonuclear Experimental Reactor) is a collaborative project between Europe, Japan, Russia and the United States. As currently planned, it would allow scientists to study the physics of a burning plasma, as well as engineering problems, such as the effect of extreme neutron bombardment on the reactor vessel walls. Results would feed into the design of a future power generating fusion reactor.

The costs, estimated to total US\$10 billion, would be shared about equally between the partners. But the collaboration is under serious strain. Earlier this month, the US Department of Energy suggested cutting its budget for ITER by three-quarters, saying it wanted to divert the money to its domestic fusion research programme (see *Nature* 391, 522; 1998). It has already sliced one-third off its budget for magnetic fusion since 1995.

European scientists have not so far been subject to such swingeing cuts. But they are aware that political enthusiasm is fragile.

Strained budgets have slowed the project's pace. Scientists accept that they will not be moving to the planned construction phase when the engineering design agreement (EDA) phase ends in July. The best they can hope is for a three-year extension to the EDA.

If the EDA is extended — which requires the agreement of all parties by July — a working group will be set up to propose cheaper options for the ITER machine. Although all partners agree that this is a good idea, they do not agree on which technical and scientific objectives could in principle be dropped.

Many US scientists now favour a more limited project that would abandon most of ITER's engineering aims, focusing instead on a purely plasma-physics programme.

That could imply a much smaller reactor whose ambitions would be confined to plasma ignition rather than sustained burn.

Most European fusion physicists are deeply opposed to such an idea, and the European Commission intends to fight to maintain an international fusion programme orientated towards the goal of a fusion energy generator. "We would not settle for a programme dissociated from energy considerations," says Klaus Pinkau, director of the Max Planck Institute for Plasma Physics in Garching, Germany.

ITER decisions are made by consensus, so a major battle between the partners seems likely. The signs in Europe are that the commission will get backing for an extension of the EDA, provided cheaper options are investigated. John Battle and Jürgen Rüttgers, the research ministers of Britain and Germany — the two European countries with the largest national fusion programmes — have expressed waning personal enthusiasm for ITER, but neither is so far pressing for the idea to be abandoned.

The extent to which European Union (EU) research ministers may be prepared to wager on fusion power solving long-term energy problems may become clearer after their meeting this week to discuss the union's fifth Framework programme of research, set to run from 1998 to 2002. Fusion research is a significant part of the programme.

Extending the EDA and maintaining national fusion programmes, nearly half of whose costs are met by the EU, would, according to the commission, require an inflation-linked continuation of the current Framework budget — ECU920 million (US\$1,013 million) over the five years. Most member states want to gamble less than this, and the EU's British presidency is aiming for a consensus between ECU770 million and ECU850 million. That would mean a drop in fusion research funding, but would not threaten an EDA extension.

Alison Abbott

...while project wins backing in Canada

[MUNICH] A Canadian consortium, ITER Canada, has been set up to promote Canada's participation in the International Thermonuclear Experimental Reactor (see above), and to push for a Canadian site to be chosen for the reactor if and when it is built. Canada's involvement is channelled through the

European partnership.

The utilities company Hydro Ontario has provided C\$6 million (US\$4 million) over the next three years, and other companies, banks and trade unions have guaranteed C\$3 million. A further C\$2 million is being sought from the federal government and the

governments of Quebec and Ontario where the two potential ITER sites are located.

"If the governments do not come in, we will have to scale down our efforts," says Peter Barnard, chairman of ITER Canada. Apart from Canada, only Japan and Italy are still offering sites.

A.A.

Drugs merger seeks strength in breadth

GLAXO WELLCOME

[LONDON] A single pharmaceutical company combining a unique breadth and depth of expertise in chemistry, biology and genomics appears likely to emerge from the merger currently under discussion between the two British-based companies Glaxo Wellcome and SmithKline Beecham.

Such a multidisciplinary approach to the search for new drugs fits in well with the previous research strategies of both companies. One traditional route to drug discovery involves choosing a 'target' disorder where a drug will be of benefit, and then experimenting with chemicals until a combination is found that will treat the target safely. This method has seen successes, such as Glaxo's anti-ulcer treatment Zantac. But there have been many more failures.

Glaxo now believes that this is too much of a 'hit or miss' approach. Under the direction of its chief executive and former director of research, Sir Richard Sykes, the company has sought to inject strategic thinking into selecting targets, finding appropriate drugs for those targets, and then using appropriate tools and skills to develop and market the drugs (see *Nature* 367, 402; 1994).

In terms of science, this has meant using cell biology, physiology and pharmacology as a guide to the cause of a disease. It means using genetics through assembling databases of patients to find genes involved in disease. And it means using combinatorial chemistry to assemble the most appropriate drug.

Combinatorial chemistry is among Glaxo Wellcome's strengths, particularly after its \$533 million purchase of the US company Affymax (see *Nature* 372, 373; 1995). Wellcome was well known for its work supporting basic biology, which the merged company continues to maintain. And SmithKline Beecham was among the first companies to invest substantially in genomics and bioinformatics.

Tim Harris, senior vice-president of research at Axy's Pharmaceuticals in California and Glaxo Wellcome's former director of biotechnology, thinks that a merger between Glaxo Wellcome and SmithKline Beecham has strong potential, on the basis of "the combination of Glaxo Wellcome's screening technology and SmithKline's genomics expertise which most other companies do not have".

Not everyone appears convinced that the new strategy will work. One delegate at a Glaxo Wellcome research seminar at the end of last year said that a sharper focus on targets and on drug design would reduce the number of early drugs failures. But he warned that the new strategy could prove costly if the company begins to lose more drugs at later — more expensive — stages of the development process.



Sykes: architect of broad-base strategy.

Jim Nidel, director of worldwide research and development at Glaxo Wellcome, said at the seminar that it was too early to judge whether the strategy is working as it had begun operating only in 1996. But he said there were several encouraging signs.

Glaxo Wellcome had introduced 28 drugs for development in the past two years, he said, of which 80 per cent were still in the pipeline. "This is better than the industry average."

"I am not claiming that we are now targeting an 80 per cent success rate; but if you remember the old success rate was 10 per cent, and we are trying to reach a 20 per cent success rate, or even 25 per cent. We are on the way to doing that."

Pharmaceutical industry estimates suggest it can cost up to \$350 million and take 10

years to bring a single drug to market. The failure rate is high. Of every 100 drugs that go into development, only ten achieve registration, and only three generate significant amounts of money.

Nidel warned that the industry would "go out of business" unless the success rate for drugs that make significant profits was increased from 3 per cent to around 10 per cent. "We will never be 100 per cent; but God love us if we cannot get the 10 per cent," he said.

Meanwhile, SmithKline has been making frequent public declaration of its confidence in its therapeutic strengths. A spokesman adds that knowledge about gene sequences and other aspects of 'genomics' is the key to future drug design.

The company struck a groundbreaking \$125-million deal with Human Genome Sciences in 1993 (see box), and subsequently appointed Peter Goodfellow — then professor of genetics at the University of Cambridge — as head of genetics research. **Ehsan Masood**

Curtain falls on gene sequencing deal

[PARIS] SmithKline Beecham, which reached a \$125-million agreement in 1993 with the gene sequencing company Human Genome Sciences (HGS), has suspended its use of that company's gene sequence database. The pharmaceutical company claims it has already exhaustively milked the database, and that other high-quality sequence data are increasingly available in the public domain.

SmithKline Beecham (SB), which is engaged in merger talks with Glaxo Wellcome (see above), has already identified more than 3,000 genes from the HGS database, according to Alain Archer, an SB spokesman. He points out that SB also recouped much of its original investment in a \$140 million deal agreed in 1996 that opened up HGS databases to Schering-Plough, Synthelabo and Takeda Industries. "SB felt that it needed to [pause] and study the targets we have already identified, rather than just searching for more," he says.

Also, says Archer, the

increasing volume of gene sequence data available in the public domain means that the debate has changed since SB signed its original agreement with HGS in 1993. In particular, this has rendered increasingly unacceptable the stringent terms imposed by HGS at the time, such as the right to royalties on products developed from sequences, and a requirement that researchers wishing to gain access to the database should give HGS an exclusive option on patents resulting from their research.

The restrictions imposed by HGS resulted in a vigorous debate about whether the human genome should be mapped in the private or public domain (see *Nature* 371, 365; 1994).

John Sulston, director of Britain's Sanger Centre, argues that this debate is now going in the direction of those who wish to see the venture being primarily public-domain. "The public databases are being refined and added to by researchers all over the world, and are of

very high quality; it is inevitable that the public domain will win under these circumstances because it becomes a much richer resource, and that is what we are now seeing."

The public-versus-private dispute "is now water under the bridge", says Richard Blevin, head of bioinformatics at Merck Sharpe and Dohme — a company that has backed public domain sequencing. Blevin argues that gaining access to sequence data is "no longer the issue" that it was a few years ago and that companies are focusing their efforts on the more difficult issue of the targets.

HGS itself has become a less prominent player in the sequencing business, following the split last year with its non-profitmaking partner, the Rockville-based Institute for Genomic Research, which has since made its sequence databases publicly available. Gene sequences are now a minor part of HGS's business, the company having focused on genome-based drugs. **Declan Butler**

Weapons physicist 'denied access' to remanufacture data

[WASHINGTON] The National Security Council intervened to block a study by the Arms Control and Disarmament Agency (ACDA) into the maintenance of the US nuclear weapons stockpile, according to the scientist who had been about to undertake the study.

Ray Kidder, a veteran nuclear weapons designer, was informed last month by ACDA that his help was no longer needed because the Department of Energy (DOE) would not give him access to a 1990 classified study of weapon remanufacture. Kidder believes that access was refused because of his public criticism of the DOE's stockpile stewardship programme — most notably an article in *Nature* last spring (see *Nature* 386, 645; 1997).

In the article, Kidder questioned the need for the stockpile stewardship programme, which uses supercomputers and major new experimental facilities to simulate nuclear weapons. He proposed a smaller alternative which would maintain the weapons by remanufacturing components as they age.

Carmen MacDougall, deputy assistant secretary for communications, says that the DOE "fully supports the need for ACDA to review the study" and that it gave the agency a technical briefing on its contents. "We offered ACDA the briefing and they accepted it," she says. "The allegations of retribution and revenge are baseless and bizarre."

The issue is sensitive for the US government. President Bill Clinton is trying to win Senate ratification for the Comprehensive Test Ban Treaty this year, on the basis of as broad a coalition of support as possible for the \$4.5 billion-a-year stockpile stewardship programme.

In November, Pierce Cordon of ACDA wrote to Kidder asking him to work for the agency as a consultant, to help determine "whether the US can be confident of weapons using remanufactured components". Cordon's letter asked Kidder to review the previous, classified study, which DOE conducted for Congress in 1990.

Kidder has the security clearance to access the study, provided he has a "need to know". A request for permission was referred to the office of Vic Reis, assistant secretary at DOE. According to Kidder, Reis could not dispute his need to know, and therefore asked Robert Bell, special assistant to the president at the National Security Council, to press ACDA to stop the study.

"My information is that Bell called ACDA and told them not to allow this to go forward," says Kidder.

But MacDougall denies that Reis asked Bell to block the study.

Colin Macilwain

Tokyo centre takes steps to boost patent output

[TOKYO] A university in Japan is aiming to increase its output of patents by turning to entrepreneurship. Its research centre is to become the country's first university organization to set up a company encouraging researchers to seek both patents on their results and private buyers for the rights to these patents.

Japanese universities, unlike their counterparts in the West, produce very few patents. According to the country's patent office, for example, although 215,400 domestic patents were filed by Japan in 1996, only seven were produced in the same year by Tokyo University.

But now this university is attempting to alter the situation. Professors from its Research Centre for Advanced Science and Technology (RCAST) have set up the Intellectual Property Company. It will be staffed by the university's postgraduate students and by staff recruited from outside. The director, too, will be appointed from outside — possibly from one of the national banks.

The company will help to evaluate the marketability and patentability of inventions made by researchers from the research centre, which specializes in areas such as biotechnology, materials science and electronics. The company will then draft patent applications and help researchers to sell the patents to industry. The last of these tasks will be carried out either directly or through intermediate agents acting for both domestic and international companies. If the sales are successful, the intermediate agents will be paid a commission. Between five and ten per cent of the proceeds from patent rights will go to the inventor, and the rest will be paid into RCAST's research budget.

Some Japanese universities have already strengthened their links with industry to promote the commercial exploitation of their inventions. Those that have already set up

university/industry liaison offices include Tsukuba University, Ritsumeikan University and Tokyo University itself. "Commercial application of research could potentially be the key to revive the Japanese economy," says one official from the Ministry of International Trade and Industry.

Until now, however, the rights to most patentable discoveries and inventions made in universities have been transferred to industry in exchange for a 'refund' of the research costs to the inventor.

Isao Karube, professor in bioelectronics and biotechnology at RCAST, and one of the founder members of the Intellectual Property Company, has invented many products, including a biosensor for O-157 E. coli. Although his research efforts have produced more than 300 potentially patentable inventions, Karube has assigned most of them to private companies for patent application.

Obtaining patents, he says, adds nothing to the status of Japanese scientists; producing scientific papers is regarded as more important. "We never have the time to file applications for patents; we often give [inventions and discoveries] away to companies."

Etsuo Niki, director of RCAST, hopes that the Intellectual Property Company will change such attitudes, as it will recruit professional staff to apply for patents. The company will also provide an opportunity for overseas companies to gain access to Japanese technology, which has long been considered inaccessible to industry outside the country, he says. "In fact, the first company to access us could well be a non-Japanese company; we are receiving a huge response from abroad," says Karube.

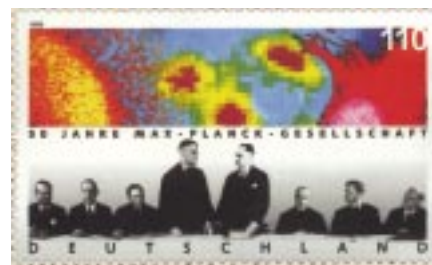
Others point out that, for new ventures such as Tokyo University's company to succeed, those involved have to demonstrate an entrepreneurial spirit, which is rare in Japanese universities.

Asako Saegusa

Stamp marks Max Planck milestone

[MUNICH] This special stamp marks the fiftieth anniversary of Germany's basic research organization. Issued last week, the stamp shows founding members of the Max Planck Society (MPS) with computer-coded images representing some of the scientific advances they pioneered: an X-ray picture of the Moon, an ion trap and a goldfish nerve.

The society was founded in 1948, despite resistance at the time from the United States to the post-fascist Germany being allowed to undertake research. It succeeded the Kaiser Wilhelm Society — itself founded in 1911



but discredited and eventually closed because of its participation in the nationalist aims of the Nazis.

Burkhardt Roeper

US Senate bills on cloning under fire from researchers

[WASHINGTON] Professional organizations representing US biomedical researchers launched a concerted attack last week on proposed cloning legislation as debate began on two bills in the Senate.

The Federation of American Societies for Experimental Biology (FASEB) described one of the bills, written by Republicans, as "damaging to worthwhile research". The Association of American Medical Colleges (AAMC) argued against any legislation, favouring a voluntary moratorium.

On 2 February, Senators Dianne Feinstein (Democrat, California) and Edward Kennedy (Democrat, Massachusetts) introduced a bill that would enact a ten-year moratorium on cloning for reproductive purposes, although allowing the cloning of human embryos that are not implanted. Breaches would be punished by civil fines.

The following day, Senators Christopher Bond (Republican, Missouri), Bill Frist (Republican, Tennessee) and Judd Gregg (Republican, New Hampshire) introduced a bill that would place a permanent ban on all forms of human cloning — including the cloning of human embryos that are not implanted. This bill bans "taking the nuclear material of a human somatic cell and incorporating it into an oocyte from which the nucleus has been removed or rendered inert and producing an embryo". Breaches would be punished by up to ten years in prison and civil fines.

In defending the Democratic bill, Feinstein argued that "the key is not to stop the technology [but] to stop the implantation of the embryo produced by this technology in a human uterus". She added that "virtually all the scientific community supports Feinstein-Kennedy and opposes Bond-Frist".

Frist argued that his bill would not impede research, but would prevent the "creating of warehouses of human embryos solely for research, and ultimately destruction".

More than 70 groups signed a letter sent by the AAMC on Monday (9 February) to senators urging them to avoid legislating, and seeking instead a voluntary five-year moratorium on research. The proposals in Congress "would have a chilling effect on vital areas of research that could prove of enormous public benefit," said the letter. It argued that the Food and Drug Administration's plan to police cloning through existing regulations on biological products "effectively protects the public" (see *Nature* 391, 318; 1998).

In a separate letter to Senator Connie Mack (Republican, Florida), the American Association for Cancer Research urged a 45-day delay on any legislative activity on cloning. **M.W.**

Population explosion raises alarm over lab animal health

[WASHINGTON] Burgeoning animal populations in US research facilities are threatening valuable animal disease models, warns the country's National Research Council (NRC). It is therefore calling for an expansion of animal holding capacity.

The NRC report *Biomedical Models and Resources* says overcrowding is posing a grave threat to the health and preservation of disease models. This overcrowding, says the report, is due to exploding populations of transgenic and knockout mice, the increased use of animals in research generally, and increased inter-institutional traffic and declining health surveillance. These factors have created "dry tinder for devastating [infectious disease outbreaks] among irreplaceable animal colonies".

As a result, the council is calling "urgently" for funding of new building to expand animal-holding capacity in research institutions, and for specialized buildings — such as the biocontainment facilities used in infectious disease research — that could be shared between researchers at different institutions.

"There's a desperate need for increased animal breeding space," says Muriel Davisson, director of genetic resources at the Jackson Laboratory in Bar Harbor, Maine, and chair of the panel that wrote the report. She urges that the fundamental need for space, diagnostic support and training should not "get lost in the glamour of specific animal models".

The report, compiled for the National Center for Research Resources (NCRR) at the National Institutes of Health (NIH), was released at a meeting of the NCRR's Advisory Council last month. Its six authors surveyed more than 70 investigators at NIH and elsewhere in compiling their recommendations, which are intended to help long-term planning at NCRR, the key supporter of animal research facilities through the NIH.

The NCRR director, Judith Vaitukaitis, calls the report — which also made recommendations on animal use ranging from training of scientists to cryopreservation of animal model embryos — "very thoughtful". She adds that the magnitude of the problem is immense and "going to get much worse".

Anecdotal evidence points to several causes for the crisis, says the report. For example, increasing mouse populations since 1980 have swollen animal numbers far beyond the level for which institutions were prepared. Also, the upgrading of standards for animal-holding facilities decreases the amount of space institutes can provide for each dollar. And animals are increasingly being shared by investigators at different institutions, so the risk of disease transmission has become



greater at a time when there are declining funds for testing for subclinical diseases. "Our laboratory animal infectious disease guard appears to be down," the report says.

Vaitukaitis says that, as a first step in responding to the situation, the NCRR is planning to establish regional centres for receiving, cryopreserving and distributing animals in the west, midwest and south of the United States. These centres would also develop specialties, for instance in models for cancer, or immunology-related research. The centres would complement the northeast's Jackson Laboratory, a unique facility that selects, cryopreserves, maintains and distributes genetically engineered mice. Such specialized centres, it is hoped, would ensure clean animal stocks.

Elsewhere, Vaitukaitis plans to focus money on training veterinarians to carry out molecular genetics, with the aim of getting them to instruct other investigators who treat animals like "four legged test-tubes", she says.

But NCRR's ability to deal with a national-scale problem remains limited, despite the hefty 13.4 per cent increase proposed in President Bill Clinton's 1999 budget plan. The centre's \$108 million 1999 budget for all work on animal disease models and facilities contains just \$7.8 million for improvement of animal facilities — merely 1.5 per cent of the total proposed spending on NCRR for 1999. Within a separate, \$20 million programme for general research construction, NCRR is seeing a steady rise in applications for funds to upgrade and enlarge animal facilities, primarily for holding knockout and transgenic mice.

Vaitukaitis says, nonetheless, that NCRR's aim remains to "significantly increase the current capacity" for animals in the nation's research facilities. Another goal, she says, is "to keep working with investigators out there to make sure we're responsive to the roadblocks that they see". **Meredith Wadman**

Germany remains split on animal testing

[MUNICH] A committee of the German parliament will this month begin work on a possible compromise between opposing factions who have been fighting over amendments to the 1986 animal protection law that were proposed last year by the cabinet.

The move follows the emergence of a new challenge to the work of German biologists, namely a proposal submitted last autumn by the opposition Social Democrat party (SPD) that the protection of animal rights should be enshrined in the constitution.

One of the main goals of the cabinet's proposal is to reduce the bureaucratic burden on scientists applying for licences for animal experiments.

The bill containing the amendments was approved by the Bundestag, the conservative-dominated lower house, last November. But the Bundesrat, the upper house in which the SPD has a majority, rejected the bill. The opposition wants stricter controls on the handling of farm animals, and a clause requiring alternatives to animal experiments to be used wherever possible.

Germany's animal protection law is the strictest in the European Union (EU). Whereas in most countries scientists holding personal licences can conduct experiments according to their own timetables, German scientists have to apply for approval for each separate experiment.

Amendment to the German law has become necessary to bring it into line with new EU rules, and the government is keen to take the opportunity to introduce other changes to make life easier for researchers. One proposed change would give automatic approval to protocols that have not been formally approved within three months of submission to regional government offices (see *Nature* 385, 760; 1997).

Researchers have welcomed this proposal, pointing out that authorities hostile to animal research often draw out approval procedures for more than six months. Scientists argue that strict regulations threaten the international competitiveness of the life sciences in Germany. "Time is a decisive factor in science, while collaboration with international research groups demands flexibility," says Jean-Alice Büttner-Enneven, a professor of neuropathology at the University of Munich.

But animal rights activists are dismayed at the proposed change. "This would be a big reversal of the achievements we have made in the past," says Jörg Styrie, scientific adviser of Bund gegen Mißbrauch der Tiere, the association opposing animal abuse.

In the current political climate, a compromise is unlikely before the federal elections in September. Nor is the proposed change to the constitution likely to be resolved soon. Demands by the SPD and the Greens that the

protection of animal rights be included in constitutional law, alongside guarantees for the freedom of science and research, would make Germany the only country in Europe where animal welfare is a constitutional issue.

Constitutional changes require a two-thirds majority in parliament, and no party or coalition is likely to achieve that in the foreseeable future. But researchers remain worried. "If animal protection became a constitutional matter, it would constantly clash with the constitutional freedom to research," says Janerik Bohling, spokesman of the Society of Health and Research, Germany's main research lobby group. And last week a local referendum voted in favour of the Bavarian government considering changes to the Bavarian constitution, including animal protection.

Academic and industrial researchers "could be driven out of Germany", says Bohling. The rights of scientists would have to be weighed up against the rights of animals during every licence approval, and challenges to the ethical status of experiments could end up in long battles in the constitutional courts.

The number of experimental animals used in Germany decreased by almost 50 per cent over the first half of the 1990s, to 1.6 million in 1995. According to the German centre for alternatives to animal testing, ZEBET, in Berlin, this is mainly due to progress in developing alternatives.

Barbara Miller

Science advice gets a face-lift in Australian reorganization

[SYDNEY] Australia's chief scientist, John Stocker, has persuaded the Coalition government to revamp its scientific advisory mechanisms, has won greater influence for scientists over policy — and has managed to have his own job upgraded.

The changes follow recommendations by Stocker in a review carried out for the former science minister, Peter McGauran (see *Nature* 385, 473 & 388, 8; 1997). The only proposal not accepted was one giving more influence to McGauran because he was not a member of cabinet. McGauran has since had to resign, and his portfolio was taken over by John Moore, already a cabinet minister.

The renamed Prime Minister's Science, Engineering and Innovation Council (PMSEIC) was originally formed under a Labor government in 1989. Stocker runs PMSEIC, which will become more powerful at the expense of the Australian Science and Technology Council (ASTEC). Stocker is responsible to both the Prime Minister, John Howard, and to Moore.

ASTEC, set up by a Coalition government 20 years ago, will be wound up later this year. On becoming chief scientist in late 1996, Stocker was made chairman of ASTEC, but



Stocker: ASTEC gives way to PMSEIC.

concluded that it was not worth continuing. He describes its two-year 'Foresight' study, intended to define national priorities, as "wasted effort producing broad generalizations that couldn't be applied". The study was quietly buried.

Announcing the changes, Howard wrote

that the emphasis of the new council "reinforces the increasingly important role the government sees science and technology playing in Australia's future". He said that the 20 members of PMSEIC (seven ministers involved in science and technology, seven representatives of academies and universities and six individuals) will be augmented by "other key representatives of the business and scientific community".

Stocker wants to include the chairs of the two major research granting bodies, the Australian Research Council and the National Health and Medical Research Council. He acknowledges that, in recent

years, meetings of the prime minister's council have been little more than "ad hoc briefings", from which no discernible changes to policy ensued.

In the new structure, the non-ministerial members will make up three working parties, taking over the role of ASTEC and charged with producing recommendations for policies and action. Stocker will chair the most important working group, on national priorities. "We aim to become a mainstream channel into policy... ASTEC could not plug into the prime minister," he says.

Topics that are politically sensitive following the government's cuts and increased charges in its first two budgets will not be avoided. The main working party's agenda includes the supply of a technologically trained workforce — science enrolments at universities have dropped markedly — and the support of basic research in universities, where infrastructure is under strain.

The full council will meet twice yearly. Stocker acknowledges pressure to deliver concrete proposals to its next meeting in mid-year, as that could be the last before a federal election.

Peter Pockley

South Korea keeps a cool head in a crisis

David Swinbanks

The economic crisis that hit Korea at the end of last year has brought hard times for the nation's scientists, but despite cuts and the effect of the plummeting won, some researchers see hope for the future.

[TOKYO] The South Korean minister of science and technology, Sook-il Kwun Kwun, surprised Western scientists at a conference last November in Seoul by declaring confidently that South Korea intended to invest \$18 billion in a biotechnology initiative over a period of 14 years ending in 2007 (see *Nature* 390, 213, 1997).

But economic crisis overtook his prediction. Within a few weeks of this statement, the Korean economy plunged into turmoil, the local currency collapsed in value, and Korea will now be lucky if it can invest half this amount in the project in dollar terms.

Throughout the crisis, which has required a massive injection of \$57 billion emergency funding arranged by the International Monetary Fund (IMF), government science officials have remained optimistic.

Despite an IMF requirement that the government should trim 10 trillion won (\$6.45 billion) — or 10 per cent — from the government budget, officials have continued to maintain that the government's budget for research and development (R&D) would be exempt from such severe reductions because of its importance to the long-term strengthening of the economy. Indeed, in December, a government panel headed by the deputy prime minister announced an ambitious plan to increase the share of government expenditure allocated to R&D from 3.9 per cent to at least 5 per cent over a five-year period (see *Nature* 391, 111; 1998).

Short-term suffering

But it is now clear that science and technology will be adversely affected, at least in the short term. According to Kwun, the budget cuts expected to be finalized shortly by the National Assembly to meet the terms of the IMF-led rescue will include a 10.2 per cent cut in the 1998 budget of his ministry, the same as the rest of the government.

Kwun still believes, however, that the 5 per cent target for 2002 is "achievable", provided that future spending on R&D increases by at least 10 per cent a year and the percentage increase for the government as a whole is no more than half this value. But in the short run, he admits, "an adjustment period might be needed".

Government researchers are already feeling the pinch. At the Korea Institute of Sci-



Taking to the streets: protests followed news of cuts, but officials hope these will be short-term.

ence and Technology (KIST) in Seoul, South Korea's first major government research institute, cuts of up to 20 per cent are being sought in direct costs of intramural research. Furthermore, industry is said to be cutting back on research contracts for government institutes, such as KIST, and universities.

"I will not be able to hire new postdoctorates or graduate students because of the high budget cut and the absence of new research contracts," says Seo Young Jeong, principal research scientist of KIST's Biomedical Research Center.

As well facing cuts in government and industry funding, Korean science is already being devastated by the decline in value of the won. Most equipment and reagents, including laboratory animals, are imported from overseas — in particular the United States — and have to be paid for in US dollars, whose value has doubled against the won.

Sunyoung Kim of Seoul National University's Institute for Molecular Biology and Genetics, for example, ordered a \$68,000 cell sorter from the United States last September. It arrived two months later, and he received the bill as the currency crisis started. "When we made the payment two weeks later, we had to pay 70 to 80 per cent more than the original price in won," he says.

Every research group in Korea is now "very sensitive" about spending dollars, says Jeong. New KIST guidelines for overseas travel are "very restrictive", leading to cuts and cancellations of travel plans, and it has

also become "very tough" to invite foreign scientists to seminars or conferences.

One initiative expected to emerge comparatively unscathed from the cuts is the Creative Research Initiative grant programme, which substantially funds some of the country's leading scientists. The initiative was launched towards the end of last year with the award of 27 large grants to individuals heading small teams of researchers in universities and government research laboratories. The initiative's new money for 1998 has been cut by only 0.6 billion won — or 4 per cent — to 14.4 billion won. But current recipients have already seen the dollar value of their grants plummet since they were awarded.

Hope of recovery

Kwun estimates that one-fifth of the ministry of science and technology's R&D expenditure — about \$140 million — will be affected by the fall in the exchange rate. But he is also hopeful that, under the IMF programme, the won should recover to a rate of 1,100–1,300 won to the US dollar from its current rate of 1,550 won (compared with a rate one year ago of 770 won).

Indeed, some leaders in the scientific community see the current crisis as an opportunity. Wan Kyoo Cho, former minister of education who is an executive trustee of the International Vaccine Institute in Seoul and head of the Bioindustry Association of Korea, says Koreans are coming to understand that science and technology are "fundamental and essential" for economic restoration.

Kwan Rim, president of the advanced institute of technology of Samsung, the giant electronics conglomerate, says his institute sees a "combination of danger and opportunity" in the crisis. "The sense of danger should be used to unite the organization around the common objective of survival, and the sense of opportunity should be capitalized on to ensure the achievement of breakthroughs."

Despite plans to restructure Samsung this year because of the crisis, Rim says the company's R&D activity will be increased. The number of personnel will increase by 9 per cent to 22,000, and the R&D budget will increase by 15 per cent, to 2.5 trillion won.

Researchers at Samsung's biomedical research institute in Seoul are being squeezed by the crisis. But, like Rim, they are confident that ambitious plans to expand research facilities at the company's adjacent medical centre, in conjunction with the company's recent establishment of a medical faculty at a private university in the outskirts of Seoul, will go ahead as planned.

Kwun is also optimistic. He says that the new administration of President Kim Dae-jung, who takes power at the end of this month, plans to restructure the government, and that the ministry of science and technology is likely to be given "greater versatility and more power".

David Swinbanks

AP/WIDEWORLD

BSE diagnosed 'more than a year before government warning'

[LONDON] The first case of bovine spongiform encephalopathy (BSE) was identified 14 months before the United Kingdom government's official announcement, according to the government pathologist who made the initial diagnosis.

Carol Richardson, a pathologist at the Central Veterinary Laboratory, told BBC television that she had identified "bovine scrapie" as the cause of death in a cow in September 1985. Despite agreeing with this diagnosis, a senior colleague attributed the death to poisoning, and failed to mention the case when reporting BSE symptoms in British cattle a year later.

Scientists at the government's neuropathogenesis laboratory in Edinburgh did not receive brain samples for testing until one year after the official announcement in November 1986.

French companies help to wire up schools

[PARIS] France is to work with computer and telephone companies to reinforce the use of the Internet and multimedia in education.

Claude Allègre, the country's minister of national education, research and technology, last week announced a series of agreements with such companies, developing a theme that emerged as the top priority in government plans, outlined last month by prime minister Lionel Jospin, to boost Internet use in France.

The flurry of deals, details of which have not yet been released, would associate the ministry with the company Alcatel for work on experimental high-speed networks, with Lotus for software for resource-sharing among groups and with Lyonnais Cable for providing cable access to schools.

Allègre said another series of agreements would be announced shortly with publishers to develop scientific and teaching materials for the national education system.

US lab 'invaded privacy' with genetics tests

[WASHINGTON] A court in the United States has ruled that a federal laboratory may have unconstitutionally invaded the privacy of employees over genetics tests carried out without their permission. The ruling, made by a regional appeals court last week, concerned seven employees who had sued the US Department of Energy's Lawrence Berkeley National Laboratory in California.

A lower court had earlier backed claims by

the laboratory that it was within its rights in testing, without permission, blood and urine samples taken from the employees before they were employed. But, in a unanimous decision, the 9th Circuit Court of Appeals overruled this interpretation and ruled in favour of the claimants — a move that appears to support demands by Vice-President Al Gore two weeks ago for legislation banning genetic discrimination in the workplace (see *Nature* 391, 429; 1998).

Hungary switches dam project back on again

[LONDON] Hungary last week revived a joint hydroelectric project on the Danube with Slovakia that it had originally agreed to 20 years ago. The Hungarian government, which had abandoned the project in 1992 under pressure from environmentalist groups, agreed in principle to complete a \$1 billion dam north of Budapest. Czechoslovakia, as it then was, completed its side of the project in 1993.

The two sides remained deadlocked for many years over Hungary's refusal to complete its share. Last year, the International Court in the Hague ordered the two sides to work out a solution that took environmental factors into account. Environment groups have accused the Hungarian government of ignoring the

international court's decision to give priority to the environmental impact of the dam project.

Indian election pledge on nuclear missiles

[NEW DELHI] India's largest opposition political party says it plans to make India an overt nuclear and missile power if it takes office after this month's elections. The election manifesto of the Bharatiya Janata Party (BJP) promises to strengthen national security by building nuclear weapons and deploying the home-made surface-to-surface missile Prithvi and the intermediate range ballistic missile Agni.

Of the Indian parliament's 544 seats, 379 are being contested by BJP and the rest by its allies in the coming elections. Although BJP cannot achieve a majority, political observers say its chances of forming a government with the help of its allies are good. The party's scientific agenda is drawn up by physicist M. G. K. Menon, a fellow of Britain's Royal Society, who joined BJP last month.

Call for UK policy review on plutonium stocks

[LONDON] Scientists in the United Kingdom have expressed concern that the country lacks a policy to deal with its stocks of 'civil

plutonium — plutonium that has been separated from reprocessed spent fuel from nuclear power stations. The Royal Society, in a report produced by a working group chaired by Sir Ronald Mason, a former chief scientist at the Ministry of Defence, urges the UK government to take steps to remedy the situation.

The report calls on the government to commission an expert review to manage the stocks. Britain currently has 54 tonnes of civil plutonium, and the stockpile is forecast to reach 100 tonnes by 2010.

Revolt in US Senate over Surgeon General's post

[WASHINGTON] Efforts by conservative Republican senators in the United States to block David Satcher's nomination for the position of Surgeon General were due to be put to the test on Tuesday (10 February), when the full Senate was scheduled to vote on the nomination.

President Bill Clinton's nomination of Satcher, the director of the Centers for Disease Control and Prevention, has been opposed by conservative Republicans led by Senator John Ashcroft (Republican, Missouri), a presidential aspirant. These Republicans criticize both Satcher's failure to oppose a particular kind of late-term abortion, and his defence of the use of

placebo in government-funded trials of the anti-AIDS drug AZT in pregnant women in Africa and Asia (see *Nature* 387, 225; 1997). The post of Surgeon General has been vacant for three years.

The American Medical Association has endorsed Satcher, as has Senator Bill Frist (Republican, Tennessee), who is a medical doctor, *The New York Times* and *The Washington Post*. Ashcroft, in a recent Senate speech, said that "Satcher's confirmation would reject America at her highest and best" and would endorse values "far beneath what the American people endorse".

Scientists put their weight behind the Gaia theory

[LONDON] Scientists met in London on Monday to launch the Gaia Society, a scientific association that aims to "promote the study of the Earth as an inter-connected living system".

The Gaia theory, which suggests that the Earth functions as a self-regulating 'superorganism', was popularized by James Lovelock, honorary visiting fellow at Green College, University of Oxford, and he is one of the scientists involved in the society. Others include Edward O. Wilson of Harvard University and Sir Crispin Tickell, chancellor of the University of Kent.

The disciplines of national wealth

Sir — Your leading article “Strategy needs disciplines” urges institutes and industries representing specific disciplines to sharpen their lobbying skills in the keen competition for funds¹. The way a country spreads its research effort across disciplines is indeed not innocent and can be gauged from its publications. A ‘publication pattern’ can even become a useful yardstick for comparing the scientific competitiveness of economies of different size.

Figure 1 shows a ‘world’ publication pattern based on the average output of the 48 most prolific countries in 18 scientific disciplines from 1981 to 1992 (a total of 6.58 million articles, proceedings, reviews and notes; calculated from data from the Institute of Scientific Information (ISI)). (The ISI database has separate entries for the Federal and Democratic Republics of Germany for the years 1981–91 and subdivides the United Kingdom into regions — England, Scotland, Wales and Northern Ireland.) In Figure 2, the further a country lies from the centre of the spiral, the more its publication pattern differs from this average pattern and the more priority it gives to specific disciplines.

England was closest to the centre, confirming the evenness of its scientific effort already noted by Sir Robert May². Close on its heels were other West European countries (Federal Republic of Germany, France, The Netherlands, Belgium, Switzerland and Italy), the United States, Israel and Canada.

This top-ten ranking does not correspond to output over the same period^{3,4} because Japan (no. 3 in output), India (no. 9) and Australia (no. 10) lie further afield, but it does include all the G7 countries (the world's largest economies) except Japan. Japan and Australia were abreast of non-G7 European countries and well ahead of the Asian ‘dragons’ which specialized more heavily.

In Southern Europe, Italy helped to set the average pattern more than Spain, Turkey or Greece, whereas in Northern Europe the Scandinavian countries formed a closely knit cluster.

In Eastern Europe, an avant-garde of two countries (German Democratic Republic and Hungary) displayed the most Western-style publication patterns although, in terms of output, they were superseded by the Soviet Union (no. 4) and Poland (no. 18, on a par with East Germany).

We have already noted the regional variations in UK publication patterns which reflect research council spending in England, Wales and Scotland⁵.

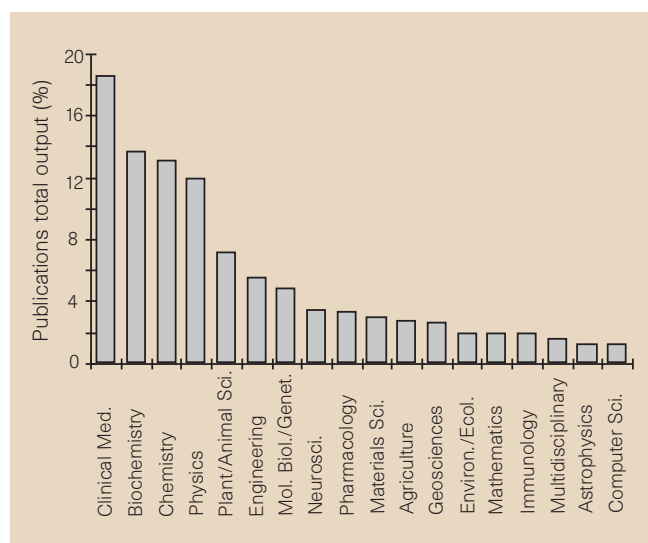


Figure 1 Average publication pattern of the 48 countries. (The disciplines are defined by ISI on the basis of journals.)

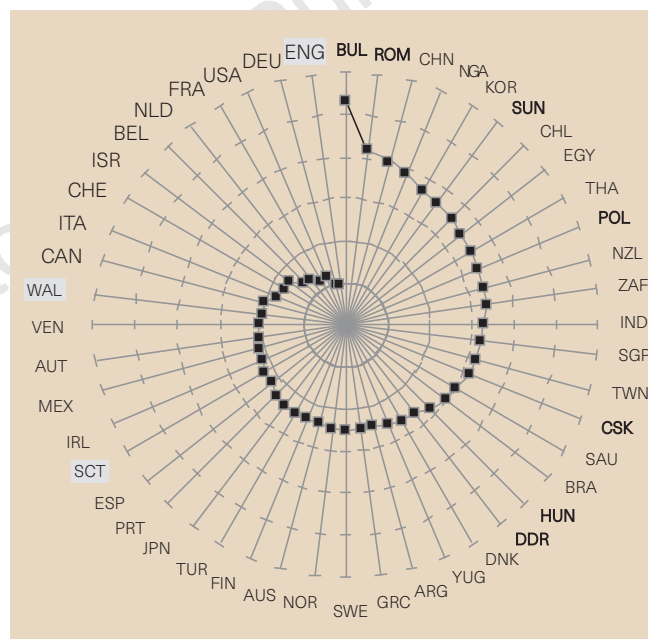


Figure 2 Chi-squared distances ($\log_{10}$ scale) of the publication patterns of the 48 countries with the highest publication outputs from the centre of gravity of the multidimensional system (48 countries $\times$ 18 disciplines). Two intervals along the radial axis are equal to 1.0. International Standards Organization country codes except for England (ENG), Wales (WAL) and Scotland (SCT). Figure adapted from ref. 3.

An average pattern depends on financial and human resources that permit research on many fronts. Excentric countries in Figure 2 may be small, suffer from an overall lack of resources motivating heavy specialization, give prominence to research based on natural resources, be governed by *dirigiste* political regimes and so on, but may also gain a leading edge in specific disciplines to become fierce competitors. Indeed, is evenness of a country's effort “a good thing”?²

Distances from the centre of gravity calculated on a yearly basis between 1981 and 1992 indicate that most countries are converging toward the centre of the spiral, that is, conforming to a world model^{6,7}. Might this increasing lack of diversity among countries not reflect increasing publication of ‘me-too’ science rather than

original research? Disciplines do not advance at the same speed and, at a specific time, it may be of strategic importance to be more productive in some disciplines than others.

The full data on which these comments are based are published in refs 2–7.

Tiiu Ojasoo

69 Alexandra Road,
Kew, Surrey TW9 2BT, UK
e-mail: ojasoo@tribunes.com

1. *Nature* **390**, 101 (1997).
2. May, R. M. *Science* **275**, 793–796 (1997).
3. Miquel, J. F., Ojasoo, T., Okubo, Y., Paul, A. & Doré, J. C. *Scientometrics* **33**, 149–167 (1995).
4. Doré, J. C., Ojasoo, T., Okubo, Y., Durand, T., Dudognon, G. & Miquel, J. F. *JASIS* **47**, 588–602 (1996).
5. Ojasoo, T., Doré, J. C. & Miquel, J. F. *Nature* **370**, 172 (1994).
6. Ojasoo, T. *Science Tribune*, <http://www.tribunes.com/tribune/art96/ojas1.htm>.
7. Okubo, Y., Doré, J. C., Ojasoo, T. & Miquel, J. F. *Scientometrics* (in the press).

Economic returns from the biosphere

Industrial companies and environmentalists are traditional opponents. But conflict may not be necessary: there is money to be made in projects that embrace environmental goals.

**Graciela Chichilnisky and
Geoffrey Heal**

Several large 'dirty' corporations, including British Petroleum, Monsanto, Dupont, Compaq, 3M, S. C. Johnson, Dow Chemical, Weyerhaeuser and Interface, are improving their financial performance by cleaning and 'greening' their operations¹. Last year, Costanza *et al.*² suggested that environmental services have great value, without indicating how this value can be realized. Here we propose various economic instruments that would allow investors to obtain economic returns from environmental assets, such as forests and landscapes, while ensuring their conservation. G. C.'s proposal³ for the creation of an international financial institution to promote this process was officially adopted by the group of 77 developing countries and by China at the Kyoto meeting last December.

The environment's services are, without doubt, valuable. The air we breathe, the water we drink and the food we eat are all available only because of services provided by the environment. How can we transform these values into income while conserving resources?

We have to 'securitize' (sell shares in the return from) 'natural capital' and environmental goods and services, and enrol market forces in their conservation. This means assigning to corporations — possibly by public-private corporate partnerships — the obligation to manage and conserve natural capital in exchange for the right to the benefits from selling the services provided. E. O. Wilson⁴ identifies "the need to draw more income from the wildlands without killing them, and so to give the invisible hand of free market economics a green thumb". Privatizing natural capital and ecosystem services is a vital step, as it enlists self-interest and the profit motive in the cause of the environment. Regulation can thus be confined to the more difficult cases.

Investing in the biosphere

In 1996, New York City invested between \$1 billion and \$1.5 billion in natural capital, in the expectation of producing cost savings of \$6 billion–\$8 billion over 10 years, giving an internal rate of return of 90–170% in a pay-back period of 4–7 years. This return is an order of magnitude higher than is usually available, particularly on relatively risk-free investments. How did this come about?

New York's water comes from a watershed in the Catskill mountains. Until recently, water purification processes by root systems and soil microorganisms, together with fil-



Attractive prospect: conservation of the environment can provide economic benefits for investors.

tration and sedimentation during its flow through the soil, were sufficient to cleanse the water to the standards required by the US Environmental Protection Agency (EPA). But sewage, fertilizer and pesticides in the soil reduced the efficacy of this process to the point where New York's water no longer met EPA standards. The city was faced with the choice of restoring the integrity of the Catskill ecosystems or of building a filtration plant at a capital cost of \$6 billion–\$8 billion, plus running costs of the order of \$300 million annually. In other words, New York had to invest in natural capital or in physical capital. Which was more attractive?

Investment in natural capital in this case meant buying land in and around the watershed so that its use could be restricted, and subsidizing the construction of better sewage treatment plants. The total cost of restoring the watershed is expected to be \$1 billion–\$1.5 billion. Hence an investment of \$1 billion–\$1.5 billion in natural capital could save an investment of \$6 billion–\$8 billion in physical capital. These calculations are conservative, as they consider only one watershed service, although watersheds (which are typically forests) often provide other important services, such as the support of biodiversity or carbon sequestration.

The commercial value of biodiversity can be partly captured by biological prospecting deals such as that between Merck and Costa Rica's InBio (see below). Joint implementation offers the possibility of commercializing

the carbon sequestration role, allowing carbon emitters in industrial countries to be credited with emission reductions that they support financially in developing countries. In other words, it allows them to buy abatement credits through bilateral trade. Several such deals have been brokered by the Global Environment Facility.

The implementation of a global multi-lateral carbon-emission market, as proposed by the United States in the context of the Kyoto negotiations, will provide a more robust way of selling sequestration services by allowing credits for carbon sequestration that can be cashed in the emissions market. In principle, then, a forest ecosystem can sell many different services. Recent provisions in Costa Rica recognize this: forested conservation areas are credited with income for the services that they provide both as watersheds and as carbon sinks, at a rate of \$50 per hectare for the former and \$10 per hectare for the latter. This is sufficient to tip the balance in favour of conserving land of marginal agricultural value.

Agriculture provides another example of the returns from investing in biodiversity to preserve genetic variation. In the early 1970s, the 'grassy stunt' virus posed a major threat to Asia's rice crop, but was defeated by the transfer of an immunity-conveying gene from wild rice to commercial varieties. In 1976, another threatening disease was defeated by transferring to commercial varieties the immunity carried by certain strains

of wild rice, preserved for just this reason by the International Rice Research Institute in the Philippines. The returns to this investment in conservation are incalculably large.

'Securitizing' the biosphere

To address its water problem (see above), New York City has floated an 'environmental bond issue', and will use the proceeds to restore the functioning of the watershed ecosystems responsible for water purification. The cost of the bond issue will be met by the savings produced: the avoidance of a capital investment of \$6 billion–\$8 billion, plus the \$300 million annual running costs of the plant. The money that would otherwise have paid for these costs will pay the interest on the bonds. New York City could have 'securitized' these savings by opening a 'watershed savings account' into which it paid a fraction of the costs avoided by not having to build and run a filtration plant. This account would then pay investors for the use of their capital.

This same financial structure is already used in securitizing the savings from increased energy efficiency in buildings. This process involves issuing contracts (securities) entitling their owners to a specified fraction of the savings. These contracts are often tradeable, issued to the providers of capital, and can be sold by them even before the savings are realized. This is a way of making investment in saving energy attractive, and does not imply any transfer of ownership of the underlying asset. The US Department of Energy has a standard protocol for estimating the savings from enhanced energy efficiency in buildings. Several financial agencies are willing to accept these estimates of energy savings as collateral for loans.

The introduction of market forces could be taken a step further. Imagine a corporation managing the restoration of New York's watershed, with the right to sell the services of the ecosystem. In this case, the service is the provision of water meeting EPA standards. Ownership of this right would enable the corporation to raise money from capital markets to meet the costs of conserving New York's watershed. If the issue was biodiversity, rather than a watershed, the corporation would own and sell (or license) the rights to intellectual property derived from the biodiversity. Such a framework would harness private capital and market forces in the service of environmental conservation.

Financing the biosphere

How significant a contribution could securitization and privatization make to conserving the biosphere? Many important watersheds are threatened by development: not only that of New York, but also the watersheds of Rio de Janeiro, the basin of the river Paraibo do Sul in the Mata Atlantica coastal forest in Brazil (a biotically unique region whose conservation would convey benefits

far in excess of the value of the water provided), and the watershed for parts of Buenos Aires. Arrangements of the type discussed here could be applied to the watersheds of some of the world's largest cities. In the United States, more than 140 cities are considering watershed conservation as an alternative to water purification. Not only could this be cost-effective, it could also stimulate conservation and a coming together of market forces with the environment.

The EPA recently estimated that ensuring safe and adequate drinking water for the United States will need infrastructure investment of \$138.4 billion over the next 20 years. The equivalent figure worldwide will be in the order of trillions of dollars. In the context of the other pressing infrastructure needs of developing countries, this amount is almost certainly not attainable by the public sector. Watershed conservation could substantially cut the investment needed, and securitization or privatization could ensure much of the balance is provided by the private sector.

What do privatization or securitization offer for other types of ecosystem? Daily⁵ identifies the following social and economic functions of ecosystem services: purification of air and water; mitigation of floods and droughts; detoxification and decomposition of wastes; generation and preservation of soils; control of most potential agricultural pests; pollination of crops and natural vegetation; dispersal of seeds; cycling of nutrients; maintenance of biodiversity; protection of coastal shores from erosion; protection from harmful ultraviolet; partial stabilization of the climate; and provision of aesthetic beauty and intellectual stimulation.

Which of these systems are amenable to the approach described here? One prerequisite is that the ecosystem must provide goods or services to which a commercial value can be attached. Watersheds satisfy this criterion: drinkable water is becoming increasingly scarce, and the availability of such water is one of the main constraints on health improvements in many poorer countries.

Commercial value of an ecosystem service is necessary but not sufficient for privatization, and some of that value has to be available for appropriation by the producer. An important issue in deciding whether ecosystem services can be privatized is the extent to which they are public goods. These are services which, if provided for one are provided for all, making it hard to exclude those who do not contribute to their costs from benefiting from their provision. So providers cannot appropriate all their returns, and for this reason we cannot be sure markets will allocate them efficiently. Water quality is a public good in the sense that if it is improved for one user it is improved for all. But water consumption is excludable, so the watershed case involves bundling a public with a private good. Knowledge, an intermediate category and one of the

services of biodiversity, has to be commercialized carefully, as shown by the need for protection such as patents and copyrights.

Ecotourism is an ecosystem service that could be treated by securitization or privatization. It is natural to expect that private investment will be forthcoming to finance the conservation of a region with significant ecotourism potential, in return for the right to some of the revenues. The growth of private game reserves is one obvious manifestation. There is a close economic resemblance to watersheds, in that the preservation of the ecosystems supporting ecotourism is a public good and benefits all. But the hotel rooms and guide services are private goods whose value is enhanced by the public good.

The commercial value of diversity is demonstrated by the International Rice Research Institute, which preserves genetic material for a range of strains (which is useful in providing immunity to new diseases, for example). Costa Rica and the pharmaceutical company Merck have made an innovative deal in which Costa Rica conserves an area of forest, supported by a payment from Merck; Merck has access to the results of biological prospecting in this forest; and Merck will pay Costa Rica a royalty on products developed from the prospecting. The deal represents a first step in providing a conservation agency in a developing country with a financial stake in the intellectual property of its biodiversity.

Can biodiversity be securitized to encourage private capital to conserve genetic variation and capture some of its commercial value? The only product of Incyte, a biotechnology company, is a database of information about genetic structures. This information has been heavily processed: biodiversity in its natural state represents unprocessed genetic information, which is less commercially usable. There may be a role for private capital in establishing a 'pre-processing' centre for genetic information from developing countries. Such a centre could conduct some preliminary analysis and sell the right to use it, with a royalty to the originating country.

For certain types of ecosystem service, privatization or securitization are real possibilities. They could be central in realizing the economic value of the underlying asset and so providing powerful economic incentives to conserve it for the future. □

Graciela Chichilnisky and Geoffrey Heal are at the Columbia University Earth Institute, 535 West 116th Street, New York, New York 10027-6902, USA.

e-mail: gmh1@columbia.edu

1. *Business Week* 10 November 1997, pp. 98–99.
2. Costanza, R. et al. *Nature* **387**, 253–260 (1997).
3. Chichilnisky, G. *Development and Global Finance: The Case for an International Bank for Environmental Settlements* (UNDP–UNESCO Office of Development Studies, New York, 1996).
4. Daily, G. (ed.). *Nature's Services: Societal Dependence on Natural Ecosystems* (Island, Washington DC, 1997).
5. Wilson, E. O. *The Diversity of Life* (Norton, 1993).

Acknowledgements. Research support from UNESCO and GEE.

Robust solutions to hard problems

John Preskill

Twenty-first century computers could achieve astonishing speed by exploiting the principles of quantum mechanics. New techniques of quantum error correction will be essential to prevent those machines from crashing.

According to the classical theory of computation, it is possible for a computing system to be fault tolerant — it can be designed so that it is sure to get the right answer even though its components occasionally fail. Now Emanuel Knill, Raymond Laflamme and Wojciech Zurek of Los Alamos National Laboratory have reported a similar result for quantum computers¹. They show that a properly designed quantum computer can achieve arbitrarily high reliability, provided that its components operate within a specified tolerance. Along with other recent developments in quantum error correction, this work has bolstered the hope that large-scale quantum

computers will someday be realized.

Whereas ordinary classical computers process information encoded in bits, a quantum computer processes information encoded in quantum states — such as the internal electronic states of individual atoms, the polarization states of photons, or the spin states of atomic nuclei. Interest in quantum computation grew explosively a few years ago when it was recognized that a quantum computer, by exploiting the exotic properties of quantum information, can make many attempts to solve a hard problem all at the same time. A quantum computer exploits a kind of massive parallelism that can never be approached by any

conceivable conventional digital computer. Therefore it can, in principle, solve certain hard problems far faster than any foreseeable digital device.

An example of a hard problem is factoring — finding the prime factors of a composite number. Nowadays, with the best hardware and algorithm, it is barely possible to find the 65-digit prime factors of a 130-digit number in a few months. The difficulty escalates sharply as the number of digits increases: with the same hardware and algorithm, we would need about the age of the Universe (10 billion years) to factor a 400-digit number. But with a quantum computer that could factor a 130-digit number in a month (which doesn't exist, of course, at least not yet), we would be able to factor a 400-digit number in just a few years, using an algorithm discovered in 1994 by Peter Shor². So at least for certain classes of hard problem, the time needed to find a solution scales much more favourably with the size of the problem if we use a quantum computer rather than a conventional computer.

This development may have profound implications for the foundations of computer science, but what of the implications for technology? When will we have quantum computers on our desktops? The hardware is still in its infancy (quantum logic gates were successfully demonstrated for the first time less than three years ago^{3,4}), so it seems safe to predict that practical and large-scale quantum computers will not be manufactured for at least several decades. But even in the longer term, one may wonder whether useful quantum computing devices will ever be feasible.

A fundamental problem is that quantum computers are far more susceptible to making errors than conventional digital computers. Complicated quantum systems are notoriously unstable; they inevitably interact with their surroundings, causing their stored information to decay — a process called decoherence. No matter how the hardware of future quantum computers is constructed, it is nearly certain that some type of error control will be needed to prevent the machines from crashing.

Error correction is a routine part of modern digital communication, but extending the classical ideas about error correction and fault tolerance to quantum devices required new ideas. The prospects for quantum computing received a tremendous boost in 1995 when Shor⁵ and Andrew Steane⁶ independently discovered that quantum error correction really is possible. If quantum information is cleverly encoded, then suitable measurements — such as those portrayed in Fig. 1 — can digitize the errors that afflict the information. If the errors are small, then most of the time the measurement projects the quantum system back to an undamaged state; rarely, the measurement projects the

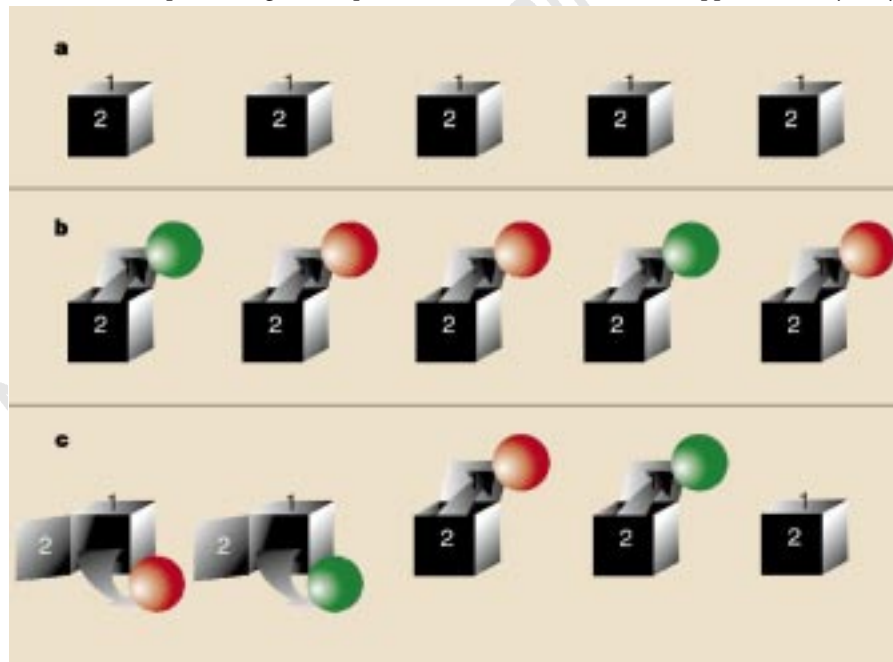


Figure 1 A simple quantum error correcting code. **a**, A quantum bit ('qubit') of information can be envisaged as a box with an object inside (here a ball). The object can be one of two colours, and the box can be opened through one of two doors; the doors might represent two different ways to measure the polarization of a photon. In the code portrayed here, one qubit of information is encoded in correlations among five different boxes. As shown in **b**, we can measure the encoded qubit by opening door 1 on all five boxes, and observing whether the number of green balls is even or odd. But to operate a quantum computer reliably, we need to be able to correct errors without measuring or otherwise disturbing the encoded qubit. An error occurs when the environment opens a door and switches the colour of the ball; different types of errors modify the correlations among the boxes in distinguishable ways. **c**, We can diagnose the error by opening door 1 of two boxes and door 2 of two other boxes — if the number of green balls is odd, then an error has been detected. Four such measurements of four boxes each suffice to identify which box is damaged and what action will repair it.

system to a state with a particular type of error that can be diagnosed and reversed. Only much more rarely is the error so serious that faithful recovery is impossible.

If we want to perform a reliable quantum computation, we must do more than merely store quantum information with high fidelity; we must also be able to process the encoded information accurately. The central principles of fault-tolerant quantum computation were formulated in 1996 by Shor⁷; now Knill, Laflamme and Zurek have invoked these ideas to show that, despite the debilitating effects of decoherence and other sources of noise, a quantum computer can carry out an arbitrarily long computation successfully, as long as the components of the machine are not too noisy. Similar results have been obtained by others^{8–10} but Knill *et al.*¹ analyse a more general and realistic class of error models.

The principles of fault tolerance can be adapted to realistic laboratory situations¹¹, and small-scale experimental demonstrations of quantum error correction are now feasible. Still, the gap between current technology and what will be needed in the future is vast, and the challenge of devising large-scale quantum computers remains formidable¹². Fresh ideas will be needed to bridge this gap. Although the new principles of fault tolerance have nourished the hope that quantum computers can overcome the menace of decoherence, a broad effort from physicists, computer scientists and engineers will be needed if quantum computers are to fulfil their destiny as the world's fastest computing devices.

If, as is indicated by efficient quantum algorithms, quantum computers can perform tasks that are beyond the grasp of foreseeable classical computers, then realizable quantum systems may have far greater potential than we now suspect to surprise, baffle and delight us. Yet this potential will never be fulfilled if we can't protect such systems from the destructive effects of noise and decoherence. Thus the discovery of fault-tolerant methods for quantum error recovery and quantum computation may have exceptionally deep implications, both for the future of experimental physics and for the future of technology. The theoretical advances have illuminated the path towards a future in which intricate quantum systems may be persuaded to do our bidding. □

John Preskill is in the Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, California 91125, USA.
e-mail: preskill@theory.caltech.edu

1. Knill, E., Laflamme, R. & Zurek, W. H. *Proc. R. Soc. Lond. A* **454**, 365–384 (1998); *Science* **279**, 342–345 (1998).
2. Shor, P. *Proc. 35th Annu. Symp. Fundamentals of Computer Science* 124–134 (IEEE Press, Los Alamitos, CA, 1994).
3. Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
4. Turchette, Q. A., Hood, C. J., Lange, W., Mabuchi, H. & Kimble, H. J. *Phys. Rev. Lett.* **75**, 4710–4713 (1995).
5. Shor, P. *Phys. Rev. A* **52**, R2493–R2496 (1995).
6. Steane, A. M. *Phys. Rev. Lett.* **77**, 793–797 (1996).
7. Shor, P. *Proc. Symp. Foundations of Computer Science preprint quant-ph/9605011 at xxx.lanl.gov*
8. Aharonov, D. & Ben-Or, M. *Proc. 29th Annu. ACM Symp. Theory of Computing preprint quant-ph/9611025 at xxx.lanl.gov*
9. Kitaev, A. Yu. *Russian Math. Surv.* No. 6 (1997).
10. Preskill, J. *Proc. R. Soc. Lond. A* **454**, 385–410 (1998).
11. Van Enk, S. J., Cirac, J. I. & Zoller, P. *Science* **279**, 205–208 (1998).
12. Preskill, J. *Proc. R. Soc. Lond. A* **454**, 469–486 (1998).

transduction. For instance, in vertebrate photoreceptor cells, the chromophore is released from opsin and regenerated in neighbouring cells; in invertebrate photoreceptor cells, the chromophore is retained and available for repeated photochemical interconversions.

This difference depends on a particular residue in the third transmembrane helix of the G-protein-coupled receptor, which in vertebrate opsins is acidic but in most invertebrate opsins (and in melanopsin) is aromatic⁴ (Fig. 1). It is likely, therefore, that melanopsin, like invertebrate opsins, is more self-sufficient in chromophore regeneration. As Provencio and colleagues point out, this feature may be important for cells such as melanophores which, because they are dispersed over much of an animal's surface, may not be able to establish a firm relationship with specialized neighbouring cells capable of providing fresh chromophore.

G proteins are used in all sorts of biological signalling systems, and some of the variety in their properties stems from differences in the amino-acid composition of their α -subunits. In the case of rhodopsin action, in most invertebrate opsins coupling is with the q-type α -subunit; in vertebrate opsins with the t (or transducin) type. One of the opsin domains implicated in this difference is the third cytoplasmic loop⁴. Melanopsin's predicted third cytoplasmic loop shows very limited sequence similarities with the corresponding loops of either vertebrate or invertebrate opsins, although its length is similar to the longer loops of invertebrate opsins (Fig. 1). This sequence dissimilarity may reflect the fact that the melanophore's pigment redistribution can be mimicked by hormones whose action is mediated through receptors that stimulate different G proteins¹. Perhaps melanopsin interacts with these G proteins and may thus allow the pigment cells to integrate the light and hormone responses.

Melanopsin's close similarities with scallop SCOP-1, and squid and octopus opsins, are not restricted to the above-mentioned functionally important domains but extend over much of the transmembrane and loop regions. They include a long cytoplasmic tail characteristic of these opsins, and clearly place melanopsin phylogenetically with invertebrate opsins. Melanopsin is thus one of the most highly divergent of a long list of opsins in vertebrates which includes other divergent opsins such as catfish parapinopsin (expressed in the parapineal and pineal gland)⁵ and vertebrate ancient opsin of salmon⁶. These opsins are also likely to be involved in non-image-forming tasks of photodetection, but are still more closely related to vertebrate ocular opsins than to melanopsin. Although convergent evolution cannot be ruled out, it can hardly

Evolutionary biology

Eyes viewed from the skin

Heinz Arnheiter

Many vertebrates and invertebrates have cells in the skin which react to light by dispersing or aggregating intracellular pigment granules. Even in a dish, such cells act like chameleons and change their shading according to the light¹.

A study by Provencio *et al.*, published last month in *Proceedings of the National Academy of Sciences*², now shows that light-sensitive pigment cells in frog skin, called melanophores, express a molecule, melanopsin, which is similar to the rhodopsins used to detect light in the eye. Melanopsin is also expressed in several other cell types known or thought to be light-sensitive in the frog, including cells in the iris, non-image-forming photoreceptor cells in the retina, and hypothalamic neurons which may be involved in controlling responses to day/night cycles. All this is hardly unexpected. Rather, the real surprise comes in Provencio and colleagues' finding that melanopsin

is more closely related to the rhodopsins in invertebrate eyes (particularly of scallop, squid and octopus) than to those in vertebrate eyes, including the frog's own eyes. This is an observation that raises several intriguing issues.

Members of the rhodopsin family are built on the principle that, when hit by photons, a chromophore related to vitamin A can flip its conformation and impart conformational changes on a seven-transmembrane G-protein-coupled receptor (opsin). Phylogenetic analyses suggest that an ancestral opsin gene was duplicated during, or even before, the split of animals into vertebrates and higher invertebrates, and then diverged to give rise to the two main branches of vertebrate and invertebrate opsins³. The main differences between members of these two opsin branches include the type of chromophore regeneration after photoactivation and the type of G-protein used for signal

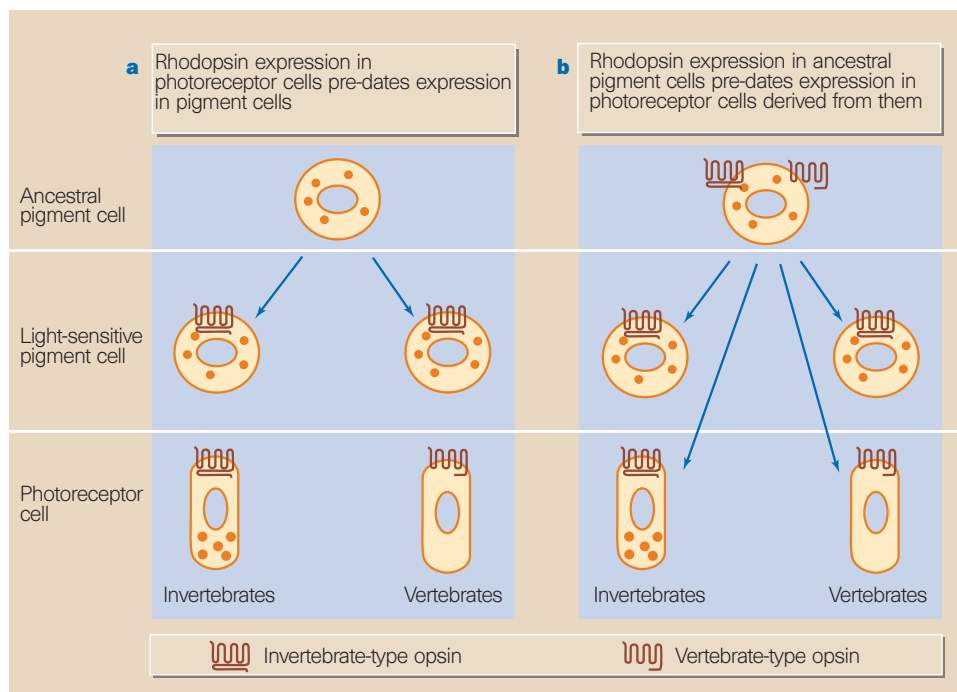


Figure 2 Two possible evolutionary explanations of the finding by Provencio *et al.*² that vertebrate melanophores express an invertebrate-type opsin. **a**, Eye photoreceptor cells and pigment cells arose independently. Gene duplications would have allowed photoreceptor cells in invertebrates and vertebrates to use different opsins, and pigment cells in both vertebrates and invertebrates may then have recruited an invertebrate-type opsin to serve their needs of light detection. **b**, Light-sensitive pigment cells of an ancestral animal containing diverged rhodopsin genes gave rise to both pigment cells and photoreceptor cells in later animals. Only vertebrate photoreceptor cells switched to vertebrate-type rhodopsins.

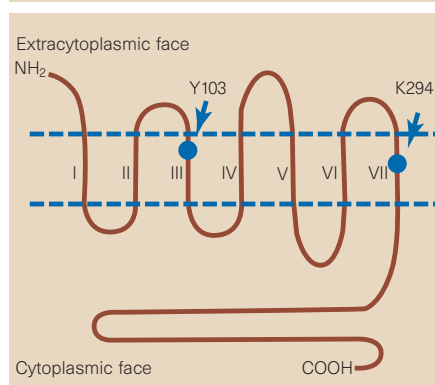


Figure 1 Opsins are seven-transmembrane G-protein-coupled receptors, and the main features of melanopsin, the form newly discovered² in the frog, are shown here. Over much of its transmembrane and loop sequence, and in the aromatic residue (Y103) which influences the type of chromophore regeneration, melanopsin is more closely related to invertebrate than to vertebrate opsins. Likewise, the long cytoplasmic tail is similar to those in scallop SCOP-1 and squid and octopus opsins. In all opsins, the conserved lysine (K294) serves as the site where the chromophore is bound.

explain the many sequence similarities of melanopsin with invertebrate opsins.

But which came first — rhodopsin expression in pigment cells or in eye photoreceptor cells? A compelling argument would seem to support the second possibility. The light response of pigment cells serves for colour adaptation, and what are colour changes good for if not for eyes to see them? Ancestral vertebrates, equipped with eyes expressing vertebrate-type opsins, may have kept invertebrate-type opsins for light sensitivity in cells outside the eye. Pigment cells

may then have recruited an invertebrate-type opsin gene for their own light-sensing needs. Recruitment of genes to function in new contexts is well documented, for instance, for lens crystallins⁷.

Alternatively, however, light-dependent intracellular pigment redistribution may have been invented not for colour adaptation but for thermoregulation and photoprotection. Because pigment cells and photoreceptor cells may share not just a rhodopsin but also pigment granules, pigment cells may in fact have been the evolutionary precursors to photoreceptor cells. The main function of pigment granules in eyes is to act as a screening agent that blocks access of light to one side of the photoreceptor cells and enables these cells to detect the direction of light.

Indeed, invertebrate photoreceptor cells often contain pigment granules of their own, and some, like those of the grass shrimp *Palaemonetes pugio*, control light scattering by redistributing the pigment granules in response to light in the same way as melanophores⁸. In vertebrate eyes in which the photoreceptor cells do not themselves contain pigment granules, the screening function is provided by neighbouring pigment cells which develop from the same tissues as photoreceptor cells. The skin pigment cells from which melanopsin was cloned are derived from the neural crest, not the neural tube from which photoreceptor cells develop; but both neural tube and neural crest are ectodermal derivatives, and so pigment cells and photoreceptor cells could still have a common evolutionary origin (Fig. 2).

If pigment cells and photoreceptor cells are indeed evolutionarily related, they may share expression of a wider array of proteins than rhodopsins, including regulatory,

structural and signalling molecules. A molecular and developmental comparison of the two cell types, particularly in invertebrates, may shed light on the question of why distinctly different eyes depend on conserved sets of molecules for embryonic development and light detection^{9,10}, yet seem to have originated in primitive eye structures that were independently invented many times¹¹. These two facts can be reconciled by the proposal that, each time a primitive eye was initiated, it started out from the same cell type, a photoreceptor cell, which provided a ready-made package of molecular components whose functions later expanded to serve the needs of auxiliary eye structures¹². A rhodopsin-expressing pigment cell in the skin of an ancestral animal would have been well poised to provide a molecular package from which eyes finally arose. □

Heinz Arnheiter is in the Laboratory of Developmental Neurogenetics, National Institute of Neurological Disorders and Stroke, Building 36, Room 5D06, 36 Convent Drive, Bethesda, Maryland 20892-4160, USA.
e-mail: ha3p@nih.gov

1. Lerner, M. R. *Trends Neurosci.* **17**, 142–146 (1994).
2. Provencio, L., Jiang, G., De Grip, W. J., Hayes, W. P. & Rollag, M. D. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
3. Kojima, D. *et al.* *J. Biol. Chem.* **272**, 22979–22982 (1997).
4. Gärtner, W. & Townner, P. *Photochem. Photobiol.* **62**, 1–16 (1995).
5. Blackshaw, S. & Snyder, S. H. *J. Neurosci.* **17**, 8083–8092 (1997).
6. Soni, B. G. & Foster, R. G. *FEBS Lett.* **406**, 279–283 (1997).
7. Piatigorsky, J. & Wistow, G. *Science* **252**, 1078–1079 (1991).
8. Doughtie, D. G. & Rao, K. R. *Cell Tissue Res.* **238**, 271–288 (1984).
9. Callaerts, P., Halder, G. & Gehring, W. J. *Annu. Rev. Neurosci.* **20**, 483–532 (1997).
10. Desplan, C. *Cell* **91**, 861–864 (1997).
11. Salvini-Plawen, L. v. & Mayr, E. in *Evolutionary Biology* Vol. 10 (eds Hecht, M. K., Steere, W. C. & Wallace, B.) 207–263 (Plenum, New York, 1977).
12. Zuker, C. S. *Science* **265**, 742–743 (1994).

Earth science

Continental mechanics

John Haines

The Earth's outer shell or lithosphere is more rigid than the immediately underlying part of the mantle. In plate tectonics, the lithosphere in oceanic regions behaves mechanically as a small number of rigid plates, with tectonic activity restricted to narrow zones where the plates meet. Continental lithosphere, by contrast, is less rigid, and deformation in continental regions is spread over hundreds to thousands of kilometres in horizontal extent. In the paper on page 655 of this issue¹, and in two others^{2,3}, England and his co-workers present results that are likely to be seen as landmarks in our understanding of the mechanical properties of continental lithosphere. They describe, respectively, observations and calculations from southern California and New Zealand, central Asia, and Greece.

In the latest paper¹, Bourne, England and Parsons report that geological slip rates on faults in southern California around Los Angeles and in the Marlborough region of New Zealand could be predicted from elastic strain accumulation measured in the past few years. They interpret the observed agreement between geologically determined long-term velocities of the crustal blocks separating the faults, and spatial averages of continuously varying, geodetically measured short-term velocities, as meaning that on both timescales the deformation of the brittle

upper crust is being driven from below by flow of the upper-mantle part of the lithosphere. The outstanding feature of the other papers, particularly that by England and Molnar², is estimation of the overall strength of the lithosphere as a strain-rate-dependent viscosity.

Central to the interpretation that flow of the subcrustal lithosphere is driving the deformation of the brittle upper crust is the assumption that the observed short-term deformation of the brittle layer mirrors steady viscous strain rates in the lower lithosphere. These strain rates cannot be measured directly. However, if the short- and long-term deformations of the upper crust were not being driven by the same, steady external forces, either from below or from the sides, there would be no reason for the velocities to be the same on both timescales.

The work of England and Molnar² is crucial to understanding that of Bourne *et al.*¹. Using an approach I pioneered for fitting continuous velocity fields⁴, England and Molnar obtained spatially compatible strain rates for the upper crust in central Asia from geological observations⁵. They approximate the velocity field as the simplest possible spline functions, obtaining spatially discontinuous strain rates as a result. But with judicious smoothing, spatial derivatives can be calculated. The real advance made by Eng-

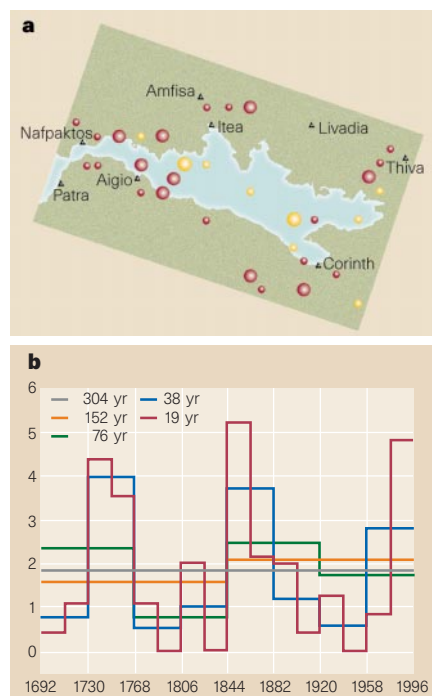


Figure 1 Seismic strain release in crustal earthquakes of $M_s \geq 6.0$ around the Gulf of Corinth, 1694–1995. **a**, Earthquakes before and after 1900 are shown as red and yellow circles respectively, and the larger circles are events of $M_s \geq 6.5$. (Reproduced from ref. 8.) **b**, Associated rate of north-south extension (units of 10^{-15} s^{-1}) inside the rectangle in different time intervals.

land and Molnar is in relating spatial derivatives of the strain rates in the upper crust to the horizontal forces associated with spatial variations of the gravitational potential

Estimating viscosity

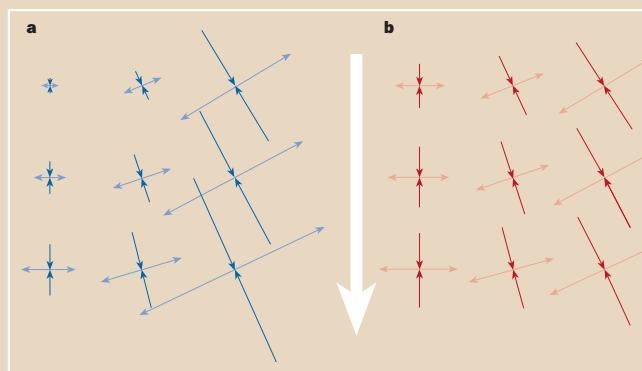
To calculate the overall viscosity of the lithosphere, England and Molnar² adopt an isotropic power-law rheology⁶:

$$\bar{\tau}_{ij} = B \dot{\epsilon}^{1/n} (\dot{\epsilon}_{ij} / \dot{\epsilon})$$

relating the depth-averaged deviatoric stresses $\bar{\tau}_{ij}$ and strain rate $\dot{\epsilon}_{ij}$ with

$$\dot{\epsilon} = \sqrt{\dot{\epsilon}_{ij} \dot{\epsilon}_{ij}}$$

$n = 3$ is appropriate⁹. The spatial derivatives of deviatoric stress which balance the gravitational forces are, therefore, dominated by spatial derivatives of either the constant B or the unit tensor of strain directions $\dot{\epsilon}_{ij} / \dot{\epsilon}$ rather than the strain rate



magnitude $\dot{\epsilon}$.

The figure here gives an example of the transformation from strain rates to deviatoric stresses, with $n = 3$. Part **a** depicts the principal strain rates, extensional (diverging arrows) and contractional (converging arrows), in a steady flow driven by a

spatially uniform force in the direction of the large arrow. Part **b** shows the associated deviatoric stresses, whose spatial derivatives counteract the driving force.

Around the margins of the Tibetan plateau there are indeed marked variations in strain directions, and England and Molnar obtain an

excellent match to the gravitational influences assuming a constant value of B . Elsewhere the match is less good, which can be explained as being due to factors resulting in variations in B .

The viscosity inferred for the strain rates in central Asia of $10^{-15} - 10^{-16} \text{ s}^{-1}$ is close to 10^{22} Pa s . The stress drop of about 10^7 Pa in continental earthquakes is a lower limit on the strength of continental upper crust. Taking this and the strain rates in central Asia gives a similar effective viscosity for the brittle layer of at least about 10^{22} Pa s . In comparison, the viscosity below the lithosphere is $10^{20} - 10^{21} \text{ Pa s}$.

J.H.

energy of the whole lithosphere⁶ resulting from the topography of Tibet. Assuming that the strain rates for the upper crust are representative of the lithosphere as a whole, this enables them to estimate the overall viscosity of the lithosphere, which they calculate as close to 10^{22} pascal seconds (Pa s) (see box and the associated figure).

In the third of the papers, Davies *et al.*³ indicate that the viscosity of continental lithosphere in the Aegean region is comparable to that in central Asia. They compare geodetic strains and earthquake occurrence in the past century in Greece, akin to a study I was involved in over a broader region of the Aegean area⁷. But with a much greater density of geodetic observations, they make the comparison on a finer spatial scale. Figure 1 illustrates the principal constraint on such analyses, using the earthquake catalogue⁸ for 300 years for an area similar in size to individual areas considered by Davies *et al.*³. Because of the random nature of earthquake occurrence in time and space, even a century can be too short to adequately estimate the long-term rates of seismic strain release in small regions.

The main contribution of England and

co-workers in these papers^{1–3} is the observations they have made relating to the mechanical properties of continental lithosphere. Their interpretations are less certain, particularly as the effective viscosity of the brittle upper crust is not less than that of the lithosphere as a whole. This neither contradicts nor confirms either their model of upper crustal deformation being driven from below, or the alternative possibility that deformation of the lithosphere is being driven from the sides by stresses that are comparable in magnitude in the upper and lower lithosphere. □

John Haines is in the Department of Earth Sciences, University of Cambridge, Bullard Laboratories, Madingley Road, Cambridge CB3 0EZ, UK.
e-mail: haines@esc.cam.ac.uk

1. Bourne, S. J., England, P. C. & Parsons, B. *Nature* **391**, 655–659 (1998).
2. England, P. & Molnar, P. *Science* **278**, 647–650 (1997).
3. Davies, R. *et al.* *J. Geophys. Res.* **102**, 24571–24588 (1997).
4. Haines, A. J. *Geophys. J. R. Astron. Soc.* **68**, 203–209 (1982).
5. England, P. & Molnar, P. *Geophys. J. Int.* **130**, 551–582 (1997).
6. England, P. & McKenzie, D. *Geophys. J. R. Astron. Soc.* **70**, 295–321 (1982).
7. Jackson, J., Haines, J. & Holt, W. *Geophys. Res. Lett.* **21**, 2849–28521 (1994).
8. Ambraseys, N. & Jackson, J. *J. Earthquake Eng.* **1**, 433–474 (1997).
9. Sonder, L. & England, P. *Earth Planet. Sci. Lett.* **77**, 81–90 (1986).

Behavioural ecology

Love thine enemy?

Bruno J. Ens

The study of cooperation and conflict between individual animals is an active and exciting field. One of the founding fathers of behavioural ecology, Nick Davies, concludes his book on dunno society¹ by saying that the lesson to be learnt is “... that we should put as much effort into our bird watching, to see how individuals behave, as we do into detailed measurements of factors influencing reproductive success”. Indeed, the comprehensive study of polygyny in the oystercatcher (*Haematopus ostralegus*), reported by Heg and van Treuren² on page 687 of this issue, does just that. This is no small achievement, given that polygyny — where one male is mated to two (or more) females — is rare in this species of bird.

Oystercatchers are typically socially as well as sexually monogamous. The male and female of a pair share parental duties such as incubation of eggs, defence of the nest and feeding the chicks. Pairs also operate as a single social unit when they compete with other pairs for territorial space (Fig. 1). But sometimes pairs break up, nearly always because the female chooses to desert her mate or because she falls victim to an usurping female³. Polygyny results when fights between females reach a stalemate. Remarkably, following the dictum “if you can’t beat them, join them”, females may then stop fighting altogether. Instead, they join forces

with the male in defending a common territory and jointly raising a brood. Why?

Heg and van Treuren addressed this question, and a large data set allowed them to rule out several hypotheses. For instance, cooperation might result from kin selection. However, evidence from microsatellite DNA convincingly showed that the cooperating

females are not more genetically related than one might expect by chance. Moreover, females don’t benefit from polygyny in terms of immediate reproduction or survival — the main benefit may be in terms of social status. Usurping non-breeding females that can’t win, but manage to stay put in a trio, are much more likely to breed in or near the territory the next year than the average non-breeding female.

It is widely believed that having two females should benefit the darwinian fitness of the male, especially when the females cooperate. Each female represents extra workforce, to help raise chicks fathered by the male. In fact, males of monogamous species have been referred to⁴ as ‘failed polygamists’. But Heg and van Treuren now show that oystercatcher males do not derive great reproductive benefits from having more mates. When nests are separate the male has to divide his assistance. With a single nest, however, the combined clutch size becomes too large for proper incubation. The groove-billed ani *Crotophaga sulcirostris* (for which joint nesting is a way of life⁵) has evolved egg tossing to help reduce its clutch size, yet oystercatchers have no such solution to the problem of incubation. This is puzzling, because other aspects of the co-operation between the females seem quite sophisticated.

One such aspect is female–female copulation. According to de Waal⁶, homosexual (and heterosexual) behaviour in bonobos is often used to reconcile two opponents after a conflict and, in this way, serves the cohesion of the group. But it will be hard to test whether homosexual behaviour in polygynous trios also promotes cooperation between the female oystercatchers. This possible social function of homosexual behaviour in oystercatchers adds to the growing



Figure 1 Two’s company — two pairs of oystercatchers contest territorial space. In their study of polygyny, Heg and van Treuren² found that an usurping female may join an existing pair to defend a common territory and jointly raise a brood.

JAN VAN DE KAM

evidence that heterosexual behaviour in the bird is not necessarily about the transfer of sperm. When oystercatcher females engage in (heterosexual) extra-pair copulations, they are thought to sample alternative mates for future breeding — not to shop around for better genes during the current breeding attempt⁷. If a female finds a 'better' partner in this way, she may then desert her current mate. Similarly, high copulation rates do not necessarily serve as a paternity guard, to dilute the sperm from competing males. Instead, well-established pairs may signal their pair bond through regular copulations, to ward off potential usurpers⁷. Thus, the oystercatcher continues to be a misfit because studies of sexual behaviour in birds are dominated by ideas derived from sexual selection and sperm competition⁸.

What incites some females to challenge others, and why do such fights sometimes end in a stalemate? And why would a male let another female in, if it does not pay to have two? Usually strange females are chased, so perhaps males are just victims of the situation. Alternatively, males may (on rare occasions) invite new females if these are of higher quality, accepting the risk of a stalemate. To answer these questions, experiments, rather than observational studies, are needed.

Such experiments would be more focused if they could test predictions derived from a quantitative theory. According to Emlen⁹, we are close to realizing the Wilsonian dream of a unified social theory. Wilson¹⁰ formulated his desire for a synthesis as follows: "... when the same parameters and quantitative theory are used to analyse both termite colonies and troops of rhesus macaques, we will have a unified science of sociobiology". However, even for a simple society such as that of the oystercatcher, there is no theory that unifies the three main career decisions for an individual — when, where and with whom to reproduce. Instead, these questions are dealt with by three largely separate fields of investigation; life-history theory, distribution theory and mate-choice theory, respectively. Whereas scientists prefer to tackle only one problem at a time, the individual bird does not have this option.

Heg and van Treuren² describe a small but interesting detail of oystercatcher society which supports my conviction that, for this species, the unifying constraint will be found in a description of how individuals acquire local dominance status¹¹. As I see it, the main difference between the oystercatcher society and the family-based, kin-structured societies that so fascinate Emlen, is that pairs — instead of families — compete for territorial space. □

Bruno J. Ens is at the Institute of Forestry and Nature Research (IBN-DLO), PO Box 167, 1790 AD Den Burg (Texel), The Netherlands.
e-mail: b.j.ens@ibn.dlo.nl

1. Davies, N. B. *Dunnock Behaviour and Social Evolution* (Oxford Univ. Press, 1992).
2. Heg, D. & van Treuren, R. *Nature* **391**, 687–691 (1998).
3. Ens, B. J., Safriel, U. N. & Harris, M. P. *Anim. Behav.* **45**, 1199–1217 (1993).
4. Birkhead, T. R. & Møller, A. P. in *Partnerships in Birds: The Study of Monogamy* (ed. Black, J. M.) 323–343 (Oxford Univ. Press, 1996).
5. Vehrencamp, S. L. *Science* **197**, 403–405 (1977).
6. de Waal, F. B. M. *Peacemaking Among Primates* (Harvard Univ. Press, Cambridge, MA, 1988).

7. Heg, D. *et al. Behaviour* **126**, 247–289 (1993).
8. Birkhead, T. R. & Møller, A. P. *Sperm Competition in Birds: Evolutionary Causes and Consequences* (Academic, London, 1992).
9. Emlen, S. T. in *Behavioural Ecology, an Evolutionary Approach* 4th edn (eds Krebs, J. R. & Davies, N. B.) 228–253 (Blackwell, Oxford, 1997).
10. Wilson, E. O. *Sociobiology: The New Synthesis* (Harvard Univ. Press, Cambridge, MA, 1975).
11. Ens, B. J., Wessing, F. J. & Drent, R. H. *Am. Nat.* **146**, 625–650 (1995).

Computational science

Unlocking dislocation secrets

Michael Marder

In the paper by Bulatov *et al.* on page 669 of this issue¹, five authors have joined forces across disciplines to study how collaboration of atoms results in change on large scales. They investigate the appearance of a barrier to plastic flow in a crystal, and then the destruction of the barrier. The larger significance of the work lies in providing a case study where complicated phenomena at the atomic scale are transformed into quantitative accounts of the whirlpools and vortices that allow metals to deform.

Vortices in fluids and gases are obvious features of everyday life, appearing when water drains from a sink, or during storms as tornadoes. The vortex has a permanent character rooted in geometry. It consists of a tight core with fluid swirling continuously about it; the core can move and become distorted, but it cannot easily be created or destroyed. J. J. Thomson, who studied the motion of vortex rings before discovering the electron, held the view that elementary particles were simply eddies of an underlying liquid ether, "matter being those portions of the liquid in which there is vortex motion"². This idea has not survived in the simple fashion Thomson had in mind. But abstract generalizations of the vortex have expanded to serve many fields of science, where they provide a compact geometrical description of complicated motions and configurations.

In the context of metals, vortices are called dislocations^{3,4}, and they solved a perplexing problem in the theory of solids. The problem lay in explaining how it is possible to bend a paper-clip. As the clip bends, the metal changes shape permanently, but all the

while remaining solid and smooth. Because the basic structure of most metals is crystalline, with each atom bound to its neighbours by strong force fields (as in Fig. 1a), it is difficult to see how the metal can flow.

Unlike the girders of a building, the bonds of Fig. 1a are not rigid but are regions of mobile charge that can snap and re-form. In 1934, Taylor⁵, Orowan⁶ and Polanyi⁷ independently realized that a vortex-like distortion of the crystal could provide the means for a wave of bond breaking and healing to move through the crystal, leaving one atom's worth of permanent distortion in its wake. The basic process is the same used by rug-layers to tighten wall-to-wall carpeting — pushing a small mound of carpet from one end of a room to the other, as shown in Fig. 1b.

The connection between vortices in liquids and solids is illustrated in Fig. 2 (overleaf). A vortex in a fluid is defined by drawing a closed loop, and summing up the fluid velocities around the loop. The result is zero if no vortex is enclosed; but if the loop circles a vortex, then the sum gives a number, κ , called the circulation. A precisely analogous procedure applies to a solid if an arrow of fixed length is assigned to every pair of nearest neighbours. Summing these arrows over a closed loop usually gives zero, but if the loop encloses the core of a dislocation, it gives a value $\vec{b}$ called the Burgers vector. The values that $\vec{b}$ can adopt are restricted to integer multiples of nearest-neighbour separations. There is no such quantization of vorticity in classical fluids, but similar restrictions apply in quantum fluids, such as

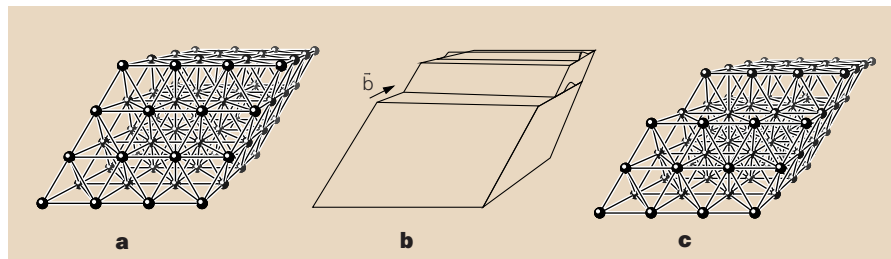


Figure 1 Metal structure and dislocations. a, The crystalline arrangement of atoms. b, The top row of atoms slides along direction $\vec{b}$ by shoving a bump from front to back; the centre of the bump is the dislocation line. c, The crystal has flowed slightly after passage of the dislocation.

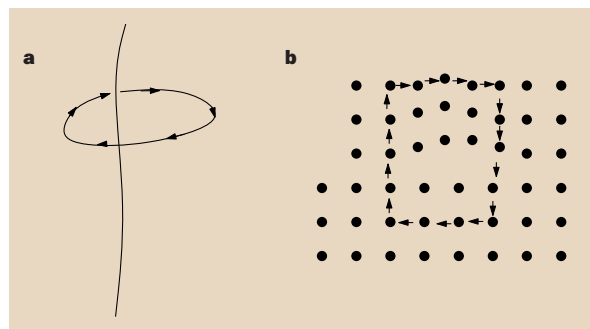


Figure 2 Vortices in fluids and solids. a, The circulation κ around a vortex in a fluid is given by summing all fluid velocities around a loop. b, The analogous quantity in a crystal is given by summing arrows between nearest neighbours over a loop surrounding a dislocation; there are four on top and three on the bottom, so they do not all sum to zero.

superfluid ^4He , or for magnetic flux lines in superconductors.

Dislocations, however, are more complicated than their fluid counterparts. The measure of dislocation vorticity, $\mathbf{b}$, is a vector. So dislocations come in several different flavours depending upon the relative direction between $\mathbf{b}$ and the dislocation line. Dislocation cores can play tricks that are unfamiliar in fluids. They sometimes decide to split into two nearby partial dislocations, each of which carries only a fraction of the quantized vorticity $\mathbf{b}$ the crystal demands, with a thin sheet of slightly defective solid connecting the two lines.

Even greater complications arise when two dislocations of differing character head towards one another and collide. Each dislocation has a particular direction in which it prefers to travel. When two combine, they have as much chance of travelling further as two dogs with their hind legs tied together on a long leash trying to run through a thick forest. Lomer⁸ and Cottrell⁹ pointed out that dislocation pairs in common metals should indeed attract each other, lock up and become immobilized; such Lomer–Cottrell locks are frequently observed experimentally¹⁰. When they are mobile, dislocations explain the ability of crystals to flow. But as they lock up, they explain the fact that the more a crystal flows, the stiffer it becomes in resisting further deformation.

The main virtue of viewing a deforming crystal as an assembly of dislocations lies in the hope of ignoring almost all detailed information on locations of the atoms and describing the evolution of the crystal purely in terms of dislocation motion. In a mesoscopic description (that is, at scales of about 10^{-6} m), dislocations are treated as curving lines that interact with one another, and with outside forces. Computer programs adopting this point of view have been written by several groups (including those of Kubin and Devincere, two of the authors of the new work¹), but have been hampered by some uncertainties. How fast should known forces make dislocations move? When they bind and interact, what force is needed to unlock them? How will the process happen?

As Bulatov *et al.*¹ show, molecular dynamics simulations, following the motion of up to 100 million atoms under the influence of Newton's laws, are finally beginning

to answer these questions. The need to follow every atom makes it impossible to study the large number of dislocations that a mesoscopic simulation can handle, but within the smaller samples that can be studied few uncertainties are left. Dislocation interactions can be studied by building numerical dislocations of various types, and measuring the forces needed to push them together or pull them apart. The results of these studies provide quantitative guidance for rules of dislocation motion in mesoscopic simulations. Ultimately, the mesoscopic simulations should provide similar guidance for the smooth, large-scale flows involved in bending something like a paper-clip.

The dislocations observed in molecular dynamics manage to play some tricks that had not fully been anticipated by decades of informed guesswork. A dislocation is seen approaching two others that have locked together. It snatches one of the dislocations

from the original pair, unzips the lock and forms a new dislocation stretching in a new direction — divorce and remarriage at the atomic scale.

Some features of the calculation are still schematic; for example, the force laws between atoms have been chosen more for computational efficiency than for maximum possible realism. However, given the interest in the geometry of dislocation motion, such details are not likely to be important, and what matters is the careful extraction of information from small scales, used as the foundation for calculations on larger scales. The ultimate goal of computer programs to mimic solids with full realism is not yet locked up, but progress is being made. □

Michael Marder is in the Center for Nonlinear Dynamics and the Department of Physics, The University of Texas at Austin, Austin, Texas 78712, USA.

e-mail: marder@chaos.ph.utexas.edu

1. Bulatov, V., Abraham, F. F., Kubin, L., Devincere, B. & Yip, S. *Nature* **391**, 669–672 (1998).
2. Thomson, J. J. *The Corpuscular Theory of Matter* (Scribner's, New York, 1907).
3. Friedel, J. *Dislocations* (Addison Wesley, Reading, MA, 1964).
4. Hirth, J. P. & Lothe, J. *Theory of Dislocations* 2nd edn (Wiley, New York, 1982).
5. Taylor, G. I. *Proc. R. Soc. Lond. A* **145**, 362–404 (1934).
6. Orowan, E. Z. *Phys.* **89**, 605–659 (1934).
7. Polanyi, M. Z. *Phys.* **89**, 660–666 (1934).
8. Lomer, W. M. *Phil. Mag.* **42**, 1327–1331 (1951).
9. Cottrell, A. H. *Phil. Mag.* **43**, 645–647 (1952).
10. Ferret, P., Robinson, B. J., Thompson, D. A. & Baribeau, J. M. *Appl. Phys. Lett.* **57**, 2220–2221 (1990).

HIV vaccines

Viral envelope fails to deliver?

Dani P. Bolognesi and Thomas J. Matthews

On 17 June 1994, the National Institutes of Health announced¹ that it would continue — but not expand — clinical trials with two vaccines against the human immunodeficiency virus (HIV), which use recombinant viral gp120 subunits to raise an immunity to the virus. To some this decision was viewed as a major setback, given the resources that had been expended in the development and testing of these vaccines. But to others, who felt that to test products with questionable promise was risky, unjustifiably costly, and would have a negative effect on future vaccine trials, this outcome was more satisfying. These events highlight a continuing struggle to establish standards for a vaccine that are acceptable to scientists, vaccine developers, government

officials and the communities affected by such trials.

Almost four years later, we have new information that raises more questions about the use of these two candidate vaccines^{2–6}. In the *Journal of Virology*, Connor *et al.*³ describe their analysis of 18 volunteers who became infected with HIV-1 while taking part in a phase II study of the two recombinant gp120 subunit vaccines. The authors searched for any indications that the vaccinated people dealt differently with their infections than did unvaccinated controls — but found none. In agreement with another study of a single case⁴, they concluded that even when the vaccine performed optimally (that is, when the full schedule of immunizations was given and

Table 1 Vaccine candidates based on primary virus envelopes

Company	Subunit	Clade	Clinical trial	Status
VaxGen	gp120	B	Phase I/II – US	Started
VaxGen	gp120	E	Phase I/II – Thailand	Imminent
Chiron	gp120	E	Phase I/II – Thailand	Imminent

A number of primary-isolate constructs are also under study, including oligomeric gp140 and gp160 constructs, pseudovirions, virus-like particles and DNA.

the recipient became infected with HIV-1 when the immune responses were near peak levels), the host responses generated had no perceptible effects on the nature or progression of the infection.

Why did these vaccines fail? One intriguing possibility, as noted by Berman *et al.*⁵, is that the subjects became infected with HIV-1 isolates bearing envelope proteins that differed from the vaccine immunogen at several neutralizing antibody-sensitive sites. But Connor *et al.*³ (and others⁶) reduced this possibility by comparing contemporary infections in matched controls. Therefore, vaccination with viral-envelope subunits alone did not exert overt selective pressure on the HIV-1 strains with which the patients became infected. Although this conclusion weakens the case for further development of these gp120 products, it is not possible to infer anything about their efficacy from these studies^{3,6} — except that the subunit vaccines are not 100% protective — because the results stem from relatively small studies.

The viral gp120 is one of two glycoproteins that make up the HIV outer envelope. It targets the virus to certain cells in the body by binding to specific cell receptors, and it initiates the early steps of virus entry into these targeted cells. Recombinant gp120 subunits emerged as candidates for vaccine development based on their ability to raise antibodies that effectively neutralize laboratory strains of HIV *in vitro*. Moreover, defined thresholds of such antibodies prevented infection of chimpanzees by the HIV strains from which the vaccines were derived. Studies for safety and immunogenicity in humans were successful, laying the foundations to test whether levels of neutralizing antibody correlated with immunity against HIV infection. But first, researchers tried to show that the neutralizing-antibody response would be effective against the viruses circulating in populations where the trial was to be carried out. In laboratory assays using field strains, however, no neutralizing activity was detectable⁷.

Why are laboratory strains of HIV sensitive to antibody neutralization, whereas primary viruses are resistant? There seem to be substantive differences between the envelope structures of these two viral phenotypes, which also use different co-receptors for infectivity. Phase I trials of several candidate vaccines that use the gp120 subunit derived from primary virus isolates are already underway (Table 1). But many researchers doubt that the simple substitution of one gp120 subunit for another will be enough to cause these constructs to act any differently from the early prototypes based on laboratory strains. Nor is there enough knowledge to design immunogens that can effectively induce neutralizing antibodies with sufficient potency and breadth against primary

viruses. Insights to this end could emerge from increased understanding of the HIV envelope, its multimeric nature, the shielding of antibody targets by carbohydrate, and the structural transitions of HIV during receptor/co-receptor binding, fusion and entry into the target cell.

The main vaccine concept currently under development is a combination of a multicomponent canarypox viral vector and one of the original gp120 envelope subunits. Such combined vaccines induce detectable cytotoxic T lymphocytes in about 60% of vaccinated people⁸. These lymphocytes can recognize and lyse target cells infected by field strains, even from distantly related virus families (clades)⁹ — in contrast to the narrow specificity of the antibodies induced by the gp120 vaccines. The antibody responses induced by the canarypox vector alone are limited, but they are substantially improved when combined with the gp120 subunit boost. Although these antibodies do not neutralize primary HIV isolates, they induce antibody-dependent cell cytotoxicity, which is detectable in 50–70% of vaccinees. Moreover, T-cell proliferative responses are more potent and durable when the two vaccines are combined⁸, and the response of CD4-positive T cells has been highlighted as a possible correlate for disease progression¹⁰.

The HIV-vaccine field remains at a crossroads. Development of an effective vaccine entails a proper balance between the growing information about HIV and empirical principles that have guided the successful production of vaccines against other agents. Correlates of protection, although useful in guiding preclinical studies, can only be established retrospectively from the results of appropriately designed clinical trials. Combining studies such as that of Connor *et al.*³ with sufficiently powerful clinical trials should allow the immunological and virological parameters that correlate with the success or failure of a vaccine to be dissected. Experiments are critical for the advancement — and, ultimately, development — of an effective vaccine against HIV. □

Dani P. Bolognesi and Thomas J. Matthews are in the Department of Surgery, Duke University Medical Center, PO Box 2926, Durham, North Carolina 27710, USA.

1. Cohen, J. *Science* **264**, 1839 (1994).
2. McElrath, J. *et al. Proc. Natl Acad. Sci. USA* **93**, 3972–3977 (1996).
3. Connor, R. I. *et al. J. Virol.* **72**, 1552–1576 (1998).
4. Kahn, J. *et al. J. Infect. Dis.* **171**, 1343–1347 (1995).
5. Berman, P. *et al. J. Infect. Dis.* **176**, 384–397 (1997).
6. Graham, B. *et al. J. Infect. Dis.* **177**, 310–319 (1998).
7. Matthews, T. *AIDS Res. Hum. Retroviruses* **10**, 631–632 (1994).
8. Crey, L. *et al. 4th Conf. Retroviruses Opportunistic Infect.* (in the press).
9. Ferrari, G. *et al. Proc. Natl Acad. Sci. USA* **94**, 1396–1401 (1997).
10. Rosenberg, E. *et al. Science* **21**, 1447–1450 (1997).



100 YEARS AGO

The War of the Worlds.

By H. G. Wells. Pp. 303.

(London: William Heinemann, 1898.)

Many writers of fiction have gathered material from the fairy-land of science, and have used it in the construction of literary fabrics, but none have done it more successfully than Mr. H. G. Wells. It is often easy to understand the cause of failure. The material may be used in such a way that there appears no connection between it and the background upon which it is seen; it may be so prominent that the threads with which it ought to harmonise are thrown into obscurity; or (and this is the worst of all) it may be employed by a writer whose knowledge of natural phenomena is not sufficient to justify his working with scientific colour. Mr. Wells makes none of these mistakes. Upon a groundwork of scientific fact, his vivid imagination and exceptional powers of description enable him to erect a structure which intellectual readers can find pleasure in contemplating.

From *Nature* 10 February 1898.

50 YEARS AGO

A new high-altitude research laboratory for cosmic ray work at a height of 11,500 ft. on the upper slopes of Monte Rosa was opened by the Italian Centre for Research in Nuclear Studies on January 11. The laboratory portion of the station at present consists of one large experimental room, and a small fully screened room for a Wilson cloud chamber. The equipment is very complete, and includes a three-phase 30 kW. power supply, a separate lighting supply, and two high-capacity battery sets with a petrol generator in case of main power failure. The station is also equipped for two-way direct radio contact with the parent laboratory at Rome, a useful facility at all times, but particularly necessary when the Laboratory is likely to be cut off from the outside world for days at a time in midwinter. ... The work at present going on in the Laboratory includes the exposure of nuclear plates, which in suitable weather can also be carried out up to 3,000 ft. above the Laboratory. Experiments on meson decay are also being carried on, and it is expected that later some Italian geneticists will be undertaking work at the station on mutations induced by cosmic rays.

From *Nature* 14 February 1948.

Photonics

More than transparent

Roy Sambles

So you thought your microwave oven couldn't radiate through the mesh of metal on the windowed door? Experiments tell us that hardly any radiation penetrates a metal plate with holes of diameter smaller than the radiation's wavelength. But it appears, from new experiments reported by Ebbesen and co-workers on page 667 of this issue¹, that such metal grids may not be as impervious to radiation as we had believed. Silver is an excellent conductor of electricity and so should screen out radiation very effectively, yet Ebbesen *et al.* found that thin, perforated silver films deposited on quartz are remarkably transparent. There is strong and selective transmission of radiation with wavelengths greater than the hole diameter.

Why should this be? Crucially, the structure is not just a random array of holes in the silver film but a regular, periodic two-dimensional grating structure (Fig. 1). The holes were 150 nm in diameter, and from 0.6 to 1.8 μm apart, in both square and hexagonal arrays. For radiation perpendicular to the film, these structures constitute zero-order diffraction gratings (for which transmission

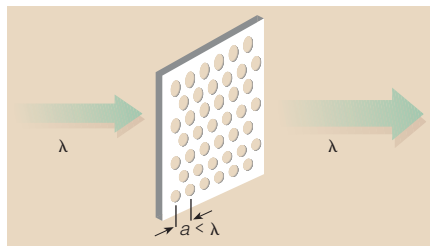


Figure 1 The phenomenon described by Ebbesen *et al.*¹. Holes in a metal screen with a diameter less than the radiation wavelength, arranged in an array with a periodicity that is also less than the radiation wavelength, can selectively transmit normal-incidence radiation.

is only possible with no deflection) for radiation of wavelength λ above 0.6 to 1.8 μm on the air side (0.88 to 2.6 μm on the quartz side). Remarkably, the 0.6- μm -spaced square-array grating transmits very strongly at 0.96 μm — with an efficiency greater than two. Increasing the thickness of the silver layer from 200 nm to 500 nm appears to increase the strength of this transmission peak.

Selective reflection has been seen before

in similar metal mono- and bi-gratings^{2,3}, but this seems to be the first observation of selective transmission. The explanation for this extraordinary behaviour appears to rest with the excitation of surface modes called surface plasmons. These are oscillating electromagnetic fields, strongly localized at the surface of a metal on which there are associated charge oscillations.

On a flat plate, an incident photon can only be converted to a surface plasmon (or vice versa) if it has the same momentum (equivalently, wave number, $1/\lambda$) parallel to the surface, and the same energy. So a normal-incidence photon cannot excite a surface plasmon. But scattering of plasmons from a periodic array of holes allows the excitation of the surface plasmon resonance, because diffraction allows the addition of multiples of the wave number to the in-plane wave number (for normal incidence this is zero).

The lowest-momentum surface plasmon (hence, lowest energy radiation, or longest wavelength) to be excited in this way has exactly the wavelength of the grid. Higher-momentum components demand higher-energy, shorter-wavelength radiation, which is readily transmitted by the holes in any case. Also, but more mysteriously, the grating geometry must somehow allow the air/silver surface modes to couple strongly with modes of different momenta on the quartz/silver interface, so energy can be transferred from one side to the other and re-radiated.

Although the theoretical analysis is incomplete, it seems clear that it is these grating-perturbed surface plasmons that allow the selective transmission of radiation. As would then be expected, the strong normal-incidence transmission occurs close in wavelength to the grating separation multiplied by the refractive index of the quartz substrate — that is, there is a small grating-induced perturbation of the surface plasmon having the highest available wavelength.

This could be a useful new way to geometrically filter electromagnetic radiation with no diffractive effects. By controlling the hole spacing, and perhaps overcoating the metal to modify the surface plasmon dispersion curve, a narrow, selectable, band of wavelengths may be transmitted over a wide angle. Combining this with voltage-controllable liquid crystals may then yield new displays or other active filter devices.

Will the phenomenon also happen at longer wavelengths? The answer might rest with the permittivity of the metal, as that dictates the momentum of the surface plasmon, at least on a planar interface. However, as the wavelength of the radiation is increased so, in general, the real part of the permittivity of most metals becomes more negative. This makes the surface-plasmon resonance sharper, and the band structure of the surface plasmons is then dictated almost entirely by the surface geometry, becoming largely

Ecology

Spring origami

Next time you pack for a scientific conference, consider how much more you can cram into a suitcase if your clothes are neatly folded. Problems of packing are common in biology — examples include DNA in chromatin, brain architecture and the arrangement of alveoli in the lung. An entertaining example from functional ecology is addressed by H. Kobayashi *et al.* in *Proceedings of the Royal Society* (265, 147–154; 1998). They have looked at the geometry of unfolding tree leaves from tightly packed buds.

The design of the leaves is reminiscent of man-made structures such as solar panels and the lightweight antennae of satellites, which have to be packed optimally to ensure fail-safe deployment. The authors have used techniques borrowed from engineering to model the leaves of two species of deciduous tree, the hornbeam *Carpinus betulus*, and the beech *Fagus sylvatica*. These species have oval, corrugated leaves with a central midrib, which develop from scaled buds.

The leaf is modelled as a plain surface with straight parallel folds. Conveniently, the creases or folds of the leaves run along lateral veins, angled at 30–50° from the



midrib of the expanded leaf. Using vector analysis, the authors calculate the effect of varying the 'vein angle' and leaf shape on packing efficiency and leaf expansion. Increasing the vein angle allows more compact folding within the bud, but only at the cost of energetically less efficient, slower leaf expansion. By getting the balance of design parameters right, natural selection may optimize the timing of leaf deployment, maximizing photosynthesis while minimizing damage caused by late frosts and leaf-eating larvae.

Rory Howlett

H. KOBAYASHI *ET AL.*

metal independent. So an appropriately periodic surface with holes smaller than the radiation wavelength should also transmit radiation selectively for almost any metal.

Fortunately for the microwave user, the mesh period and the hole diameter are so much smaller than the radiation wavelength that there is no chance of leakage through this process. □

Roy Sambles is in the Department of Physics, Stocker Road, University of Exeter, Exeter EX4 4QL, UK.

e-mail: j.r.sambles@exeter.ac.uk

1. Ebbesen, T. W., Lezec, H. J., Ghaemi, H. F., Thio, T. & Wolff, P. A. *Nature* **391**, 667–669 (1998).
2. Wood, E. L., Sambles, J. R., Cotter, N. P. K. & Kitson, S. C. *J. Mod. Opt.* **42**, 1343–1350 (1995).
3. Watts, R. A., Harris, J. B., Hibbins, A. P., Preist, T. W. & Sambles, J. R. *J. Mod. Opt.* **43**, 1351–1360 (1996).

Immunology

Unmasking the killer's accomplice

Marco Colonna

Class I molecules of the major histocompatibility complex (MHC) are polymorphic glycoproteins which, under normal circumstances, are expressed on the surface of almost every cell in the body. Upon viral infection or in tumour cells, however, MHC class I can be downregulated, and its absence is 'sensed' by natural killer (NK) cells in a process known as recognition of 'missing self'¹. This is mediated by cell-surface receptors which, on binding class I molecules, transduce inhibitory signals that block NK-cell-mediated lysis. When class I expression is lost or reduced, NK cells are released from inhibition and can, therefore, rapidly identify and eliminate virally infected or tumour cells.

Intriguingly, some MHC class I NK-cell-receptor homologues promote — rather than inhibit — the activation and cytotoxicity of NK cells, by a hitherto unknown mechanism. But on page 703 of this issue, Lanier *et al.*² describe the key to this mystery. They have discovered a signal-transducing molecule called DAP12, which couples an MHC class I receptor on the surface of the NK cell to an activating signal-transduction pathway within that cell.

Natural killer cells monitor the expression of class I molecules using many cell-surface receptors. Molecular cloning of these receptors has revealed remarkable diversity^{3,4}. Two receptor families have been identified: the C-type lectin molecules (including rodent Ly49 receptors and heterodimers of the human CD94 and NKG2 polypeptides); and molecules belonging to the immunoglobulin superfamily, known as human killer-cell inhibitory receptors (KIR).

Each receptor family includes many members that recognize different class I molecules. Heterogeneity has also been observed in the cytoplasmic domains of the receptors. Inhibitory receptors have a long cytoplasmic tail that contains immunoreceptor tyrosine-based inhibitory motifs (ITIMs). On recognition of class I ligands, ITIMs are phosphorylated, allowing recruitment and activation of a protein tyrosine phosphatase⁵. This results in dephosphorylation events that block NK-cell activation

and, thus, the cytotoxic response. But, in each family, many NK-cell receptors (such as Ly49D, CD94/NKG2C and KIR2DS) have short cytoplasmic domains that lack ITIMs and do not inhibit the cytotoxic response. In fact, these truncated receptors transmit stimulatory signals to NK cells by activating protein tyrosine kinases and phospholipase C, leading to an increase in NK-cell-mediated cytotoxicity^{6–9}.

How do these receptors initiate NK-cell activation on binding class I molecules? Because activating receptors lack the sequence motifs that are implicated in positive signal transduction, they probably transduce signals through associated proteins. Use of a separate subunit to mediate signal transduction is not unusual among immunoreceptors — FcεRI and FcγRIII, for example, are oligomeric complexes in which signal transduction is mediated by associated protein subunits. The T- and B-cell receptors also use separate subunits, and all contain one or more immunoreceptor tyrosine-based activation motifs (ITAMs) in their cytoplasmic domain¹⁰. When the ligand-binding component of the receptor is engaged, the cytoplasmic ITAMs are tyrosine-phosphorylated by src-family tyrosine kinases. This leads to the recruitment of ZAP-70 and SYK protein tyrosine kinases¹⁰, which trigger a cascade of intracellular phosphorylations that lead to cellular activation.

Biochemical studies¹¹ have shown that one activating NK-cell receptor is noncovalently associated with a protein (relative molecular mass, M_r 12,000) that exists as a disulphide-linked dimer. Following a logical and insightful path, Lanier and colleagues² have now identified the complementary DNA that encodes such a molecule. They screened a database for cDNAs encoding unidentified ITAM-bearing proteins, and pulled out DAP12, which encodes a transmembrane protein (M_r 12,000) that strikingly resembles the γ -chain of the FcεRI and the ζ -chain of the T-cell receptor. The cytoplasmic tail of DAP12 contains a single ITAM that is tyrosine phosphorylated and can bind in this form to both ZAP-70 and SYK tyrosine kinases (Fig. 1).

The DAP12 protein associates with one activating receptor, KIR2DS2. The resulting complex, when expressed in pre-B or T cells, induces events that lead to cell activation. Moreover, in the hydrophobic transmembrane region of DAP12 is a negatively charged aspartic-acid residue that is thought to interact with the positively charged lysine residue in the transmembrane domain of the class I-binding subunit (Fig. 1). Finally, DAP12 has a very short extracellular domain that contains cysteine residues which, presumably, mediate interchain disulphide linkage.

Interestingly, DAP12 is expressed not only in NK cells, but also in other types of cell that are involved in the immune response, such as dendritic cells. So could DAP12 also be involved in the activation of these cells? Receptors bearing transmembrane and cytoplasmic domains similar to those in activating NK-cell receptors have been identified in myeloid and lymphoid cells¹², representing potential partners for DAP12. The murine pre-T-cell receptor also associates with a disulphide-linked dimer (M_r 12,000) that may correspond to DAP12 or to a related molecule¹³.

Although the discovery of DAP12 is a breakthrough in our understanding of the mechanism by which activating MHC class I receptors transduce stimulatory signals, the biological function of these receptors is still

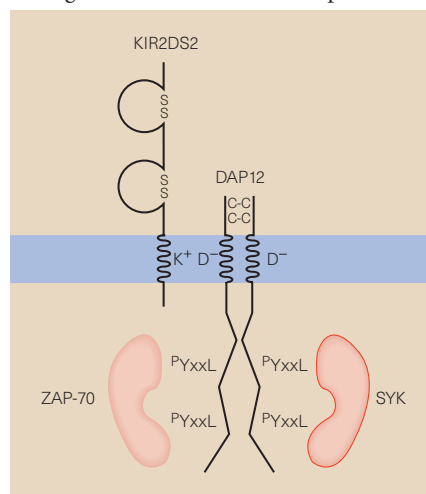


Figure 1 Structure of the activating natural killer (NK)-cell receptor complex for major histocompatibility complex (MHC) class I molecules. The complex consists of a class I-binding component, the KIR2DS2 receptor, associated with the DAP12 disulphide-linked homodimer identified by Lanier *et al.*². DAP12 contains a cytoplasmic immunoreceptor tyrosine-based activation motif that is tyrosine phosphorylated upon receptor ligation, and recruits ZAP-70 and SYK protein tyrosine kinases. These kinases trigger a cascade of phosphorylation events that lead to cellular activation. The KIR2DS2 receptor and DAP12 are non-covalently associated — complex formation is facilitated by interaction of transmembrane residues with opposite charges.

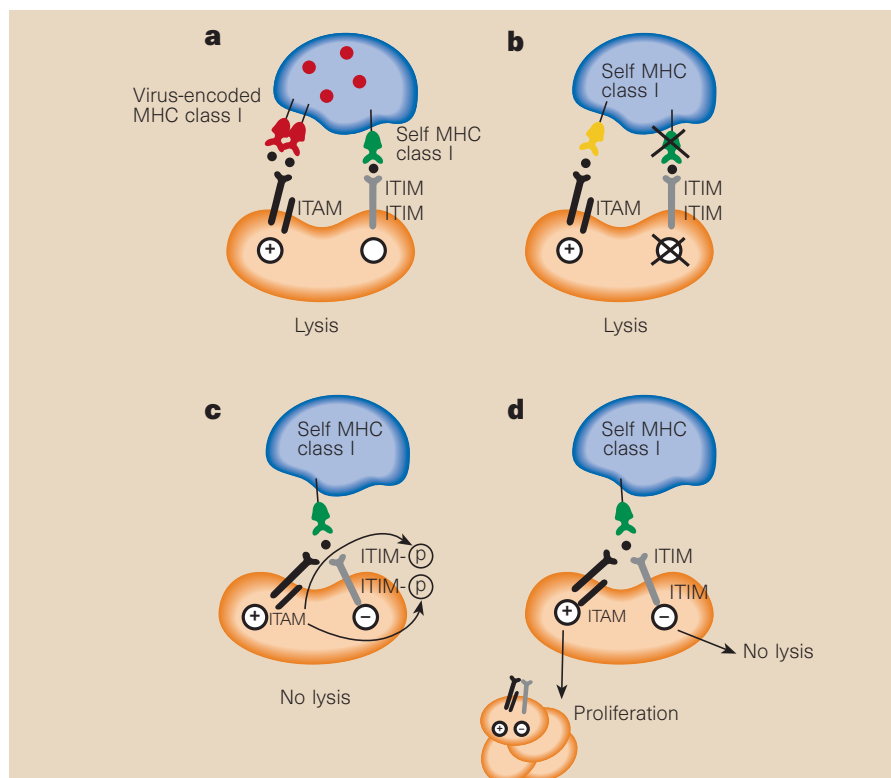


Figure 2 Hypothetical interactions between activating and inhibitory MHC class I receptors. Inhibitory and activating receptors may recognize self and non-self class I molecules, respectively (a), or distinct self-class I molecules (b) on the same cell. Activating receptors may allow NK cells to recognize target cells that express virally encoded MHC class I molecules/ self class I molecules carrying viral peptides (a), or target cells that have selectively lost one class I molecule (b). Alternatively, inhibitory and activating receptors may recognize the same class I molecules on a target cell. In this case, activating receptors may regulate inhibitory receptors by phosphorylating cytoplasmic immunoreceptor tyrosine-based inhibitory motifs (ITIMs; c), or may control distinct NK-cell functions. Activating receptors may promote NK-cell proliferation, whereas inhibitory receptors may control effector responses such as cytotoxicity (d).

enigmatic. How can inhibitory and activating MHC class I receptors interact to regulate the activity of NK cells? Does the existence of activating receptors violate the 'missing self' model? One hypothesis is that inhibitory receptors recognize self MHC class I molecules (preventing autoreactivity), whereas activating receptors detect non-self MHC class I molecules, such as those encoded by cytomegalovirus to deceive NK-cell inhibitory receptors¹⁴ (Fig. 2a). Another possibility is that individual NK-cell clones express both activating and inhibitory receptors that are specific for different self-MHC class I ligands. These receptors may cooperate to identify class I deficient cells. For example, the preferential loss of a class I ligand specific for an inhibitory receptor could allow activation of the NK cell through an activating receptor, which recognizes another class I ligand on the target cell (Fig. 2b). Alternatively, if both inhibitory and activating receptors see the same MHC class I molecule on a given target cell, the activating receptor may cooperate with the inhibitory receptor by recruiting tyrosine kinases to phosphorylate the cytoplasmic ITIMs (Fig. 2c). Finally, inhibitory receptors may regu-

late effector and cytotoxic responses, whereas activating receptors could mainly control NK-cell proliferation, promoting NK-cell expansion during development (Fig. 2d). Now that Lanier *et al.*² have discovered the mechanism by which the activating receptors signal, it will be easier to define the biological function of activating receptors in NK-cell recognition of tumours and virally infected cells. □

Marco Colonna is at the Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland.
e-mail: colonna@bii.ch

1. Ijünggren, H. G. & Kärre, K. *Immunol. Today* **11**, 237–244 (1990).
2. Lanier, L. L., Corliss, B. C., Wu, J., Leong, C. & Phillips, J. H. *Nature* **391**, 703–707 (1998).
3. Yokoyama, W. M. *Curr. Biol.* **5**, 982–985 (1995).
4. Lanier, L. L. *Immunity* **6**, 371–378 (1997).
5. Leibson, P. J. *Immunity* **6**, 655–661 (1997).
6. Mason, L. H. *et al. J. Exp. Med.* **184**, 2119–2128 (1996).
7. Houchins, J. P., Lanier, L. L., Niemi, E., Phillips, J. H. & Ryan, J. C. *J. Immunol.* **158**, 3603–3609 (1997).
8. Moretta, A. *et al. J. Exp. Med.* **182**, 875–884 (1995).
9. Biassoni, R. *et al. J. Exp. Med.* **183**, 645–650 (1996).
10. Weiss, A. & Littman, D. R. *Cell* **76**, 263–274 (1994).
11. Olcese, L. *et al. J. Immunol.* **158**, 5083–5086 (1997).
12. Samaridis, J. & Colonna, M. *Eur. J. Immunol.* **27**, 660–665 (1996).
13. Takase, K. *et al. J. Immunol.* **159**, 741–747 (1997).
14. Kärre, K. & Welsh, R. M. *Nature* **386**, 446–447 (1997).

Daedalus

Craters of doom

The Earth is under heavenly threat. Every so often, a comet or asteroid hits it with sufficient violence to bring about a mass extinction of higher species. Many people advocate a nervous watch for such celestial missiles, and the building of huge nuclear-armed rockets to deflect them.

Yet, oddly, recent mass extinctions seem nonrandom, having occurred at about 26-million-year intervals. One theory gives the Sun a companion star ('Nemesis'), in a highly eccentric orbit with just this period. Every time Nemesis swings past the Sun, she disturbs the Oort comet cloud, and may bring comets of her own into the Solar System. In the resulting storm of comets, several hit the Earth. If even one is moderately big, disaster follows.

So far, Nemesis has proved elusive, but Daedalus plans an indirect check. The Moon must also come under cometary attack; and craters on the Moon are not smoothed away by weather. So he advocates examining the Moon for impact craters whose ages are multiples of 26 million years. If found, they would confirm the theory, and allow us to predict the next likely attack.

But how to tell the age of a lunar crater? Daedalus notes that the Moon's surface is steadily eroded by meteoritic dust impacts, whose debris forms a layer of 'regolith'. A meteoritic or cometary impact would excavate the regolith at the impact site, replacing it with characteristic 'new crater' surface. Subsequent dust impacts would slowly cover this surface deeper and deeper in standard regolith.

The pristine crater surface and regolithic debris must differ in thermal capacity and conductivity. These can be measured by the rate of cooling when sunlight is cut off by a lunar eclipse. Daedalus wants to do this continuously, by infrared and microwave imaging of the dark Moon's surface just beyond the 'terminator' defining local sunset. The longer wavelengths of this thermal emission will have come from deeper layers of the surface. As the terminator sweeps across the lunar disc, the changing temperature profile behind it should disclose recent craters. The wavelength-dependence of their thermal emission will show their depths of regolith, and hence their ages. Statistical study should reveal the period of Nemesis, and when she is due to strike next. With luck, the results will set our minds at rest. If we turn out to have (say) 13 million years in hand, the nuclear defence programme will lose its urgency.

David Jones

Obituary

Reginald Victor Jones (1911–97)

Physicist who was instrumental in Britain's war effort

R. V. Jones, who died on 17 December 1997, was from 1940 to 1945 director of scientific intelligence for the Air Ministry and then for the British Intelligence Service. His grasp of physics enabled him to develop new systems and battle tactics that influenced the course of the Second World War, while his whimsical, sophisticated but highly intelligent 'double think' baffled the German High Command.

Jones was born on 29 September 1911. His father was a sergeant with long service in the Grenadier Guards, and after the First World War the family grew up at Aldershot. At Alleyn's School in south London, Jones was a member of the officer training unit, which reinforced his background in military tradition. Then, at Wadham College, Oxford, he gained first class honours in physics finals. After his doctorate in 1936–37, he was elected a Skynner Student in astronomy at Balliol College.

I first met Jones in 1937 when I started research on low-temperature physics in the old Clarendon Laboratory. I took over from George Pickard, who had just started working with Jones on research into infrared physics in the attics there. On several occasions they went together to the Royal Aircraft Establishment at Farnborough; on 27 April 1937, Jones managed to detect another aircraft in flight, probably the first time this was ever done with infrared radiation.

By a well-established custom, all the senior members and research students met for morning coffee and afternoon tea in the central area of the Clarendon, which was surrounded by large glass-fronted cabinets containing scientific equipment. These still exist in the new Clarendon Laboratory, to which we moved just as war broke out in September 1939. The head of the laboratory, F. A. Lindemann, was a close friend of Winston Churchill, becoming his chief scientific adviser in 1939 and continuing throughout the war. He was later ennobled as Viscount Cherwell.

Another well-known physicist at the laboratory was Derek Jackson, who measured the hyperfine structure of the caesium atom, and deduced its nuclear magnetic moment by estimating the distance of the electron from the nucleus.



Of this he said: "Sommerfeld did it by wave mechanics; I did it by arithmetic". In 1942 Jackson joined the Royal Air Force, becoming an observer and gunner on night fighters, and scoring many successes. Also at the Clarendon were A. H. Cooke (who later invented the 'T. R. Tube', which made possible single-aerial working for radar on aircraft), and J. H. E. Griffiths (who in 1945 discovered ferromagnetic resonance).

Jones had moved over to war work in 1937, and made many crucial contributions in the desperate times to come. After he had devised the airborne infrared set to detect the heat from aircraft engines, Henry Tizard, then chairman of the Aeronautical Research Committee, and rector of Imperial College, suggested that Jones should move to Imperial to continue his investigations into infrared detectors. They were thought of as an alternative to radar. The final verdict was to concentrate on radar, although Lindemann remained worried that the enemy might find a way of jamming it. His fears stemmed from a suggestion made by Jones that this could be done by dropping from aircraft metal strips of an appropriate size, a device that was later known by the code-name 'window' and was indeed used.

In the summer of 1940, evidence from various sources convinced Jones that the enemy were using radio beams ('Knickebein') to guide their aircraft. He put this to Lindemann, who took him to present the evidence to Churchill in person; countermeasures were introduced just in time. On another technological front, in 1942 the Germans introduced a device to detect metre-wave radar from British aircraft, giving their submarines time to escape attack by submerging.

When radar at centimetre waves was proposed for raids over Germany, Jones protested that its first use should be at sea. He reasoned that, when a bomber was shot down over land, the central and innovative component, a magnetron, would soon fall into enemy hands. His advice was not taken, but fortunately the Germans did not realize that a similar device was being used at sea.

At the end of hostilities, Jones became professor of natural philosophy at the University of Aberdeen, Scotland. His lectures were greatly appreciated by undergraduates for the numerous demonstrations, which took much time to prepare. On the research side, Jones investigated the limits of measurement through delicate instrument design. He also created a unit growing single crystals of magnetic materials, which he made freely available for experiments elsewhere, including the Clarendon.

From Aberdeen, he was recalled three times to reconstruct the British scientific intelligence service that he had created and headed during the war. His account of those times, *Most Secret War* (1978; published in the United States as *The Wizard War*), was a cool, detailed account of the secret technological battles in which he played a leading role. He struggled against disbelievers among the armed forces, "who rejected the logic of analysis", and against "the misdirection, misinterpretation and fragmentation of technical information". Jones described the importance of anticipating the enemy's use of science in warfare in his article "Scientific intelligence", published in *Research* (9, 347–352; 1956).

In 1982 General Doyle Larson, head of the Electronic Security Command of the United States Air Force, invited Jones to lecture to his command. Jones became an honorary member of the USAF and honorary mayor of San Antonio in Texas. Two years later, at the headquarters of the Central Intelligence Agency in Langley, Virginia, the first R. V. Jones Intelligence Award was presented, naturally enough, to Professor Reginald Victor Jones. In 1993, he became a Companion of Honour, one of the highest civilian distinctions in Britain.

Jones married Vera Cain in 1940; they lived together in Aberdeen until she died some 50 years later, just after the death of their eldest daughter, Susan. Jones stayed on in his university house after his retirement, even when the roof began to collapse.

Brebis Bleaney

Brebis Bleaney is at the Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.



Wright's *An Eruption of Vesuvius, Seen From Portici*, c. 1774–76.

Wright's ructions

A volcanic eruption was a fitting subject for Joseph Wright. The artist had links with some of the great scientific thinkers in an age when science was beginning to discover the secrets of the forces that shape the Earth.

Martin Kemp

Joseph Wright of Derby's continued residence in his home town, far from condemning him to provincial isolation, located him within the remarkable orbit of the Lunar Society, whose luminaries included Josiah Wedgwood, James Watt, Matthew Boulton, Erasmus Darwin and Joseph Priestley.

Wright's exceptional range of subject-matter extends beyond conventional landscapes and portraits into iron forges, industrial premises, alchemists' dens, air pumps and orreries, and plunges us visually into the rising scientific ferment of the second half of the eighteenth century. The sciences of chemistry, not least discoveries relating to combustion and oxygen, and of physics, particularly electricity, were revealing new dimensions to the forces that animate nature.

Volcanology played a significant role in demonstrating that the vast formative forces in the body of the Earth lay far beyond the scope of the biblical creation and Flood in time and magnitude. The discipline may be said to have been founded by Sir William Hamilton, King's Envoy in Naples, antiquar-

ian and passionate observer of geological phenomena. In Hamilton's view, the now fertile fields of lava must be "so very ancient, as to be far out of the reach of history". Rugged topographies resulted not from God's static design but from huge effusions of material molten by subterranean fires — "mountains are produced by volcanoes, not volcanoes by mountains" — a process that was still very much under way. Indeed, Hamilton believed that the present surface of the Earth contains no surviving part from its original, 'primitive' state.

Wright visited Vesuvius in 1774, two years after Hamilton had published his *Observations on Mount Vesuvius....* The painter wrote excitedly back to John Whitehirst in Derby that he had witnessed "a very considerable eruption... of which I am going to make a picture". Four years later Whitehirst was to publish his *Inquiry into the Original State and Formation of the Earth*, which talked about the great sea of inner fire that surged through what Wright called the "bowels of the mountain". In the event, Wright painted not one picture of Vesuvius, but more than 30.

His series of pictures, composed in the stu-

dio as sublime exercises in visual poetry, are seemingly different in method from Hamilton's professed system of "collecting facts".

The vertical jet of white-hot debris, the vortex of clouds and emergent Moon correspond less to what Wright witnessed at a single moment than to an artistic synthesis of the essence of Vesuvius's grandeur within the drama of world history.

The granular globs of white paint that comprise the jet and bedeck the fiery slopes stand for, rather than literally describe, the phenomenon.

Yet Wright's vision does strike to the very heart of the vitalistic and animistic enthusiasms that did so much to convince European scientists such as Luigi Galvani, Antoine Lavoisier, Alexander von Humboldt and Humphry Davy that the great key to the mystery of life lay in the kinds of universal electrochemical and combusive forces which twitched a frog's leg into life and breathed "Oxygenated Air" into every creature. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Sex and violence in hermaphrodites

Differences in objectives between males and females are a driving force in the evolution of copulatory mechanisms. Hermaphrodites can also have sexual conflicts, but caused by opposing sexual interests within rather than between individuals. One consequence seems to be that physically damaging sex, as occurs with the marine flatworm *Pseudoceros bifurcus*, might be favoured more in hermaphrodites than in species with separate sexes.

Most animals transfer sperm *in copula*, a genital 'handshake' under strong female control¹. However, a widespread alternative involves injecting sperm under the partner's skin. This method is well-known in hermaphrodites such as leeches², flatworms³ and sea slugs⁴, but rare when sexes are separate¹. For males, injecting sperm offers direct access to eggs, whereas females bear the costs of wound healing and lose control over fertilization¹. Selection on females to avoid these costs must be strong but, in a hermaphroditic population, individuals can weigh the costs of being stabbed against the benefits of stabbing others. Observations on *Pseudoceros bifurcus* Prudhoe, a polyclad, marine flatworm, show that individuals indeed attempt to increase sperm donation over sperm receipt.

We collected *P. bifurcus* at Heron Island (Queensland, Australia) at 15–20 metres depth. Seventeen pairs were assembled such that each individual could be recognized by size or coloration. We observed them in 4-litre containers with seawater and coral sand. After a first session (20 hours, seven pairs) we found mating activity did not change with time, so we shortened later sessions to 6 h (five pairs) and 4 h (five pairs). Sexually interacting individuals rear up and evert their penis on contact (Fig. 1a). When both individuals rear up, escalated contests ensue, lasting for 20.3 ± 12.6 minutes (mean $\pm$ s.d.) (maximum 59 min; $n = 39$, 14 pairs). Individuals try to stab one another, but show strong avoidance behaviour when struck by their partner. During 39 contests, 287 strikes led to 46 inseminations in 12 pairs. Cases where only one animal reared up were frequent ($n = 72$) but resulted in only four inseminations because the non-responding individual glided out of reach.

These results suggest 'rearing up' has evolved specifically to interact with a sexually responding partner. The behaviour is effective for 'striking and parrying': the average number of inseminations per contest was similar for both partners within a pair (intraclass correlation coefficient⁶ $r_i = 0.65$, $F_{(11,12)} = 4.6$, $P = 0.007$), indicating that the chances of hitting or being hit were closely matched. Importantly, most inseminations were unilateral ($n = 30$ events; Fig. 1b). Even when reciprocal penis insertion could be achieved by the second partner ($n = 10$ events), the first to inseminate obtained a longer injection time than the second (9.5 ± 2.9 and 5.7 ± 4.2 min, respectively; paired $t_{(6)} = 3.04$, $P = 0.023$; $n = 7$ pairs; data weighted per pair). The asymmetrical outcome of inseminations favours animals that inject first, as they will father more eggs in more partners and have fewer wounds to heal (Fig. 1c).

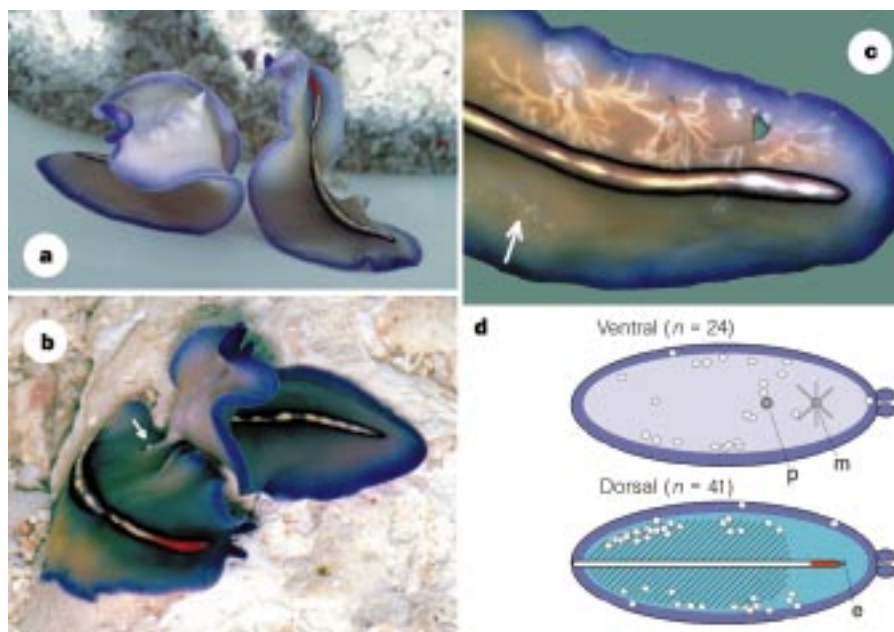


Figure 1 'Penis fencing' in *P. bifurcus*. **a**, Two individuals (4–6 cm long) rearing up with everted penises, attempting to inseminate their opponent while avoiding being inseminated. The left individual is turning away, having been struck unsuccessfully. **b**, Unilateral insemination by one partner. Sperm is transferred for the duration of penis insertion, resulting in a sperm droplet under the partner's skin (arrow). **c**, Sperm moving through the body after several inseminations. The two holes in the upper half were caused by ventral and dorsal inseminations in the same place. Time since insemination: 24–36 h for the upper half, 3 days for the white dots (arrow). **d**, Summary of all injection sites (17 experimental pairs and exploratory observations). Ovaries are many and small, scattered dorsally (shaded area)⁵, making ventral injections less effective, as the widely branched gut separates ventral and dorsal halves. p, penis; m, mouth; e, light-receptor cells.

Although 'penis fencing' can best be explained as an attempt to increase the benefit of sperm donation over the cost of sperm receipt, the exact relation between insemination and fertilization is still unknown. Avoidance of insemination might not only cut costs but also offer the benefit of being inseminated by better 'stabbers', resulting in more successful offspring. Mate choice alone, however, seems insufficient to explain the evolution of a costly and risky behaviour in which the winner is determined primarily by chance and where sperm receipt seems to be in the wrong place in many cases (Fig. 1d). Since animals often do not meet again for hours, repeated contests between the same two partners must be rare in the wild.

The mating system of *P. bifurcus* differs

from that of hermaphrodites without hypodermic insemination, where partners cannot take advantage of each other and typically exchange gametes by controlled reciprocity (ref. 7 and refs therein). Hypodermic insemination, when present, allows hermaphrodites to skew sexual interactions in favour of sperm donation, fuelling an evolutionary arms race between strike and avoidance behaviour.

N. K. Michiels

Max-Planck-Institut für Verhaltensphysiologie,
Seewiesen, Postfach 1564,
D-82305 Starnberg, Germany
e-mail: michiels@mpi-seewiesen.mpg.de

L. J. Newman*

Zoology Department, University of Queensland,
QLD Australia 4072

1. Eberhard, W. G. *Female Control: Sexual Selection by Cryptic Female Choice* (Princeton Univ. Press, Princeton, 1996).
2. Malecha, J. in *Reproductive Biology of Invertebrates 5 (Sexual Differentiation and Behaviour)* (ed. Adiyodi, K. G. & Adiyodi, R. G.) 231–248 (John Wiley, Chichester, 1992).
3. Apelt, G. *Mar. Biol.* **4**, 267–325 (1969).
4. Tompa, A. S., Verdonk, N. H. & van den Biggelaar, J. A. M. *The Mollusca 7 (Reproduction)* (Academic, Orlando, 1984).
5. Newman, L. J. & Cannon, L. R. G. *Mem. Queensl. Mus.* **37**(1), 205–266 (1994).
6. Sokal, R. R. & Rohlf, F. J. in *Biometry* 213–214 (Freeman, New York, 1995).
7. Petersen, C. W. *Envir. Biol. Fishes* **43**, 351–361 (1995).

* Present address: National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA.

Gas hydrate crystals may help build reefs

During a recent cruise in the Porcupine Basin, off southwest Ireland, we discovered two extensive and hitherto largely unsuspected deep-water reef provinces, including a giant cluster of hundreds of buried mounds. The ring shapes of many reefs suggest that they are caused by an axial fluid expulsion at the sea bed, a transient flow well confined in space and time. We are exploring various hypotheses, but a stimulating avenue for research is opened by a glacially controlled growth pulse and subsequent decay of a shallow layer of gas hydrates as a methane buffer and probably indirectly as a ground for overlying biological communities.

Seismic profiles in the Porcupine Basin had already identified a number of large deep-water carbonate mounds. These are up to 2 km in diameter, with heights of 250 m (ref. 1). Gravity cores on the mounds recovered deep-water colonial coral debris (*Lophelia* sp.), identical to the fauna collected by Wyville Thomson in the area during HMS *Porcupine*'s cruise in 1869 (ref. 2). A high-resolution seismic survey with R/V *Belgica* in May 1997 revealed that these large isolated seabed mounds are only one of several reef types in the basin. The large mounds are flanked to the north by a 1,200 km² province of smaller reefs, 50–100 m in height, covered by a few tens of metres of sediments

(Fig. 1a, b). These are the Magellan reefs, named after the survey vessel that had observed them a few months earlier. The large surface reefs seem to be associated with fault-controlled methane seeps from deeper hydrocarbon reservoirs¹, but the numerous buried reefs, rising from, or rooted to, a common and virtually undisturbed geological horizon, do not show any obvious correlation with deeper faults. Reflection seismograms of the Magellan reefs reveal the widespread twin associations of symmetric lung- or butterfly-shaped structures, suggesting central cross-sections through ring structures (Fig. 1a, c).

Such ring reefs call for an axial causal mechanism. Singular methane vents or mud volcanoes could have been girdled with 'stone rings' of reef-building communities. There is still a poor understanding³ about whether chemosynthetic and symbiotic communities^{4,5} that could contribute to reef-building processes benefit directly from the seeping gases, or whether deep-sea reef construction in seep environments is simply a matter of preferential settlement and growth on solid, authigenic carbonates, which have been widely reported as primary deposits fringing seep sites⁶.

The sudden onset and near-synchronous decay of the well-confined, giant Magellan mounds province with its numerous ring reefs implies a regional and transient degassing event, not fully compatible with the pace of deeply rooted geological processes. Here a palaeoclimatic control can be invoked. Environmental and spatial argu-

ments indicate the possible mediating and buffering role of a shallow methane hydrate reservoir, which might have grown and spread widely upslope from discrete deep feeder pipes, particularly in glacial times, when the Porcupine sea bed was swept by colder bottom currents. The upslope boundary of the Magellan province, where the reef structures fade out, follows depth contours of the sea bed (Fig. 1b), supporting the idea of a pressure control on the extension of the hydrates. When such a hydrate layer, usually capping a reservoir of free gas, subsequently decayed and cracked in warmer postglacial waters, it could have funnelled methane over hundreds of seeps to the sea bed.

Gas hydrates are ice-like crystalline compounds that occur when water molecules form a cage structure around guest molecules under conditions of high pressure and low temperature, which are mostly found in the marine environment below water depths greater than 500 m (ref. 7). Their generation generally requires the presence of a prolific methane source. The occurrence of gas hydrates on various margins of the oceans, mostly associated with hydrocarbon seeps, is being increasingly demonstrated; their relevance to world margin stability, the marine biosphere and climate is of significant topical relevance. Although other possible models are being explored, the methane hydrate mediation hypothesis, if verified, would shed new light on the interaction between submarine gas seeps and a climate-controlled interplay between seabed gas hydrates and the deep ocean biosphere.

Still another, fundamentally different, reef setting has been discovered on the eastern margin of the Porcupine Basin: huge barrier-like reefs with ponded sediments, rising from an enigmatic, acoustically transparent and deeply eroded sole layer, which seem to fringe a buried, fossil shelf edge at water depths of about 750 m (Fig. 1e). This is reminiscent of Nordic settings⁸ of small cold-water reefs on shelf breaks, where hydrodynamic conditions (internal waves) and nutrient fluxes shape preferred sites for suspension-feeding communities.

Such discoveries indicate that the Porcupine Basin is a major deep-water coral reef province, while highlighting some hitherto unsuspected habitats of reef-building communities in deep, cold waters. They provide a potential hydrocarbon indicator in a petroliferous basin in the North Atlantic, illustrate a range of deep-water carbonate habitats and provide clues to the Late Tertiary and recent evolution of the basin.

J. P. Henriët, B. De Mol, S. Pillen, M. Vanneste, D. Van Rooij, W. Versteeg
Renard Centre of Marine Geology,
University of Gent, Krijgslaan 281,
9000 Gent, Belgium

P. F. Croker
Petroleum Affairs Division, DTEC, Beggar's Bush,

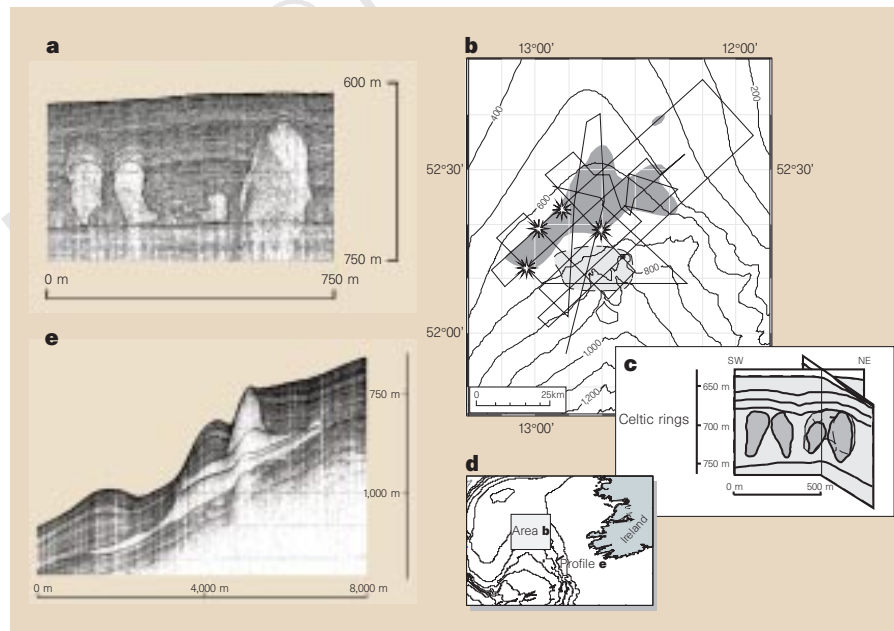


Figure 1 Magellan reefs. **a**, Illustrative seismogram of a central cross-section through a small buried ring reef (left) and a slant section through a larger ring reef, buried below about 30 m of sediments. **b**, R/V *Belgica* track map and outline of the Magellan reef province (black), north of the occurrence of large surface mounds (area circled by dotted line). A few buried reefs that reach the surface are indicated by stars. Bathymetry is contoured in metres. **c**, Line drawing of cross-sections of ring reefs on intersecting profiles. **d**, Location of map (**b**) and profile (**e**). **e**, The barrier-like 'Belgica reefs' fringing a buried shelf edge on the eastern slope of the Porcupine Basin.

Haddington Road, Dublin 4, Ireland

P. M. Shannon, V. Unnithan

Department of Geology, University College Dublin,
Belfield, Dublin 4, Ireland

S. Bouriak, P. Chachkine

UNESCO Centre for Marine Geology
and Geophysics,

Moscow State University, Vorobjevy Gory,
119899 Moscow, Russia

The Porcupine–Belgica 97 Shipboard Party

e-mail: jeanpierre.henriet@rug.ac.be

1. Hovland, M., Croker, P. F. & Martin, M. *Mar. Petrol. Geol.* **11**, 232–246 (1994).
2. Thomson, C. W. *The Depths of the Sea* (Macmillan, London, 1873).
3. Sassen, R., Cole, G. A., Droz, R. & Roberts, H. H. *Mar. Petrol. Geol.* **11**, 55–65 (1994).
4. Fisher, C. S. *Rev. Aquat. Sci.* **2**, 399–436 (1990).
5. Hovland, M. & Judd, A. G. *Seabed Pockmarks and Seepages* (Graham & Trotman, London, 1988).
6. Reilly, Jr, J. F., MacDonald, I. R., Biegert, E. K. & Brooks, J. M. *AAPG Memoir* **66**, 39–62 (1996).
7. Sloan, E. D. *Clathrate Hydrates of Natural Gases* (Marcel Dekker, New York, 1997).
8. Henrich, R. *et al. Facies* **32**, 71–108 (1995).

Role for lichen melanins in uranium remediation

Lichens are successful colonizers in extreme terrestrial habitats world-wide, including metalliferous environments¹. Their ability to accumulate metals has led to their use in monitoring radionuclide fall-out from Chernobyl and uranium uptake from dust resulting from mining². Here we report for the first time a lichen growing directly on uranium minerals and uranium being concentrated within its tissues. Our study suggests that melanin-like pigments, substances previously unreported within lichens, are involved.

We discovered the lichen *Trapelia involuta* growing directly on the secondary uranium minerals metazeunerite ($\text{Cu}(\text{UO}_2)_2(\text{AsO}_4)_2 \cdot 8\text{H}_2\text{O}$) and metatorbernite ($\text{Cu}(\text{UO}_2)_2(\text{PO}_4)_2 \cdot 8\text{H}_2\text{O}$) on uraniumiferous spoil heaps in Cornwall³. X-ray element maps for uranium across transverse sections through the lichen–rock interface show uranium accumulation principally within the outer fruiting body (apothecia) walls of the lichen compared with their interior (Fig. 1a, d). These high uranium concentrations are not caused by the trapping of mineral particles, as reported in previous studies, because U-to-Cu, U-to-As and U-to-P ratios within the lichen differ significantly from those of possible trapped mineral phases. Complexing with oxalic acid and secondary lichen products can also be discounted as mechanisms because we identified calcium but no uranium-bearing oxalates within the apothecial walls. We irrigated sections with the oxidizing agent sodium hypochlorite, confirming the presence of the secondary metabolite

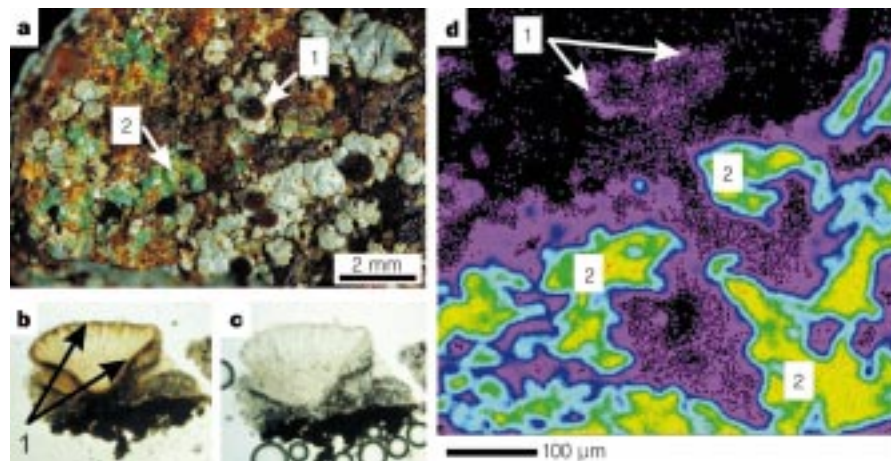


Figure 1 Lichen concentrating uranium. **a**, *Trapelia involuta* (1) with dark brown apothecia growing on metazeunerite ($\text{Cu}(\text{UO}_2)_2(\text{AsO}_4)_2 \cdot 8\text{H}_2\text{O}$) (2); **b**, transverse section of *T. involuta* apothecia showing pigmented wall, without treatment with sodium hypochlorite; **c**, the same section with wall decolorized by irrigation with sodium hypochlorite; **d**, X-ray element map showing uranium accumulation (1) in apothecia wall of *T. involuta* growing on metazeunerite (2).

gryphoric acid within the main lichen body, but not within the apothecia, suggesting that this is also not responsible for the uranium accumulation.

Trapelia specimens present on uraniumiferous spoil have atypically dark brown–black (rather than pale red–brown⁴) apothecia, the pigment being restricted to their outer walls and visible in thin section (Fig. 1). Melanin is known to form in non-lichenized fungal hyphae as a response to a wide range of environmental stresses, including metal contamination⁵. We therefore extracted the pigment⁶ from about 100 *Trapelia* apothecia. Fourier transform infrared spectra of the dark pigment indicated the presence of aromatic organic compounds similar to humic acids⁷. This, together with the observation that the dark pigment is decolorized by sodium hypochlorite (Fig. 1b, c), suggests that the compound is melanin or melanin-like. In the samples studied, there is a strong correlation between the localization of melanin and high concentrations of uranium, which agrees with experimental studies⁶ showing that melanin has a high adsorption capacity for uranium. It is possible that melanization in *T. involuta* is an adaptive response to protect the ascospores, the sexual reproductive bodies formed within the apothecia, from the toxic effects of the uranium.

Our findings are an important step towards enhancing the use of lichens as biological monitors of radionuclides and potentially in remediation processes. Laboratory studies on yeast and fungal biomass have shown effective uptake of uranium, leading to the biological treatment of metal-contaminated effluents. In a pilot study⁸, pellets of the non-lichenized fungus *Aspergillus niger* were used successfully in a fluidized-bed reactor for the removal of uranium, a process 14 times more effective than a commercial ion-exchange resin. The ability of lichens to

concentrate metals from solution is also demonstrated by the development of a biosorbent based on immobilized lichens⁹. Because lichens grow slowly, it is unrealistic to expect them to be used directly in bio-remediation programmes to concentrate metals. However, mechanisms of bioaccumulation are poorly understood in microorganisms, and lichenized fungi have an important role. Specific lichen tissues have been shown to concentrate secondary metabolites to more than 20% dry weight¹⁰. Localization and characterization of organic metal compounds in organisms are a major challenge for analysts; the capacity for lichens to accumulate such high concentrations makes them model systems for studying fungus–metal interactions.

Other lichen species in uraniumiferous environments must now be studied, to allow us to assess the wider response of lichens to uranium. Other species might concentrate uranium to far higher concentrations and by different mechanisms. By identifying how lichens accumulate uranium, remediation technologies based on these mechanisms might be developed.

J. McLean, O. W. Purvis, B. J. Williamson

The Natural History Museum, Cromwell Road,
London SW7 5BD, UK

e-mail: w.purvis@nhm.ac.uk

E. H. Bailey

Department of Physiology and
Environmental Science, University of Nottingham,
Sutton Bonington, Leicestershire LE12 5RD, UK

1. Purvis, O. W. *Sci. Prog.* **79**, 283–309 (1996).
2. Fahselt, D., Wu, T.-W., & Mott, B. *Bryologist* **98**, 228–234 (1995).
3. Dines, H. G. *The Metalliferous Mining Region of SW England* (Mem. Geol. Surv. Gt Br., London, 1956).
4. Purvis, O. W., Coppins, B. J., Hawksworth, D. L., James, P. W. & Moore, D. M. *Lichen Flora of Great Britain and Ireland* (Natural History Museum, London, 1992).
5. Gadd, G. M. *New Phytol.* **124**, 25–60 (1993).
6. Ellis, D. H. & Griffiths, D. A. *Can. J. Microbiol.* **20**, 1379–1386 (1974).

7. Filip, Z., Haider, K., Beutelspacher, H. & Martin, J. P. *Geoderma* 11, 37–52 (1974).
8. Yakubu, N. A. & Dudeney, A. W. L. in *Immobilisation of Ions by Bio-sorption* (eds Eccles, H. & Hunt, S.) 183–200 (Ellis Horwood, Chichester, 1986).
9. Ramelow, G. J. et al. *Int. J. Environ. Anal. Chem.* 53, 213–232 (1993).
10. Huneck, S. & Yoshimura, S. *Identification of Lichen Substances* (Springer, Berlin, 1997).

Electric current stimulates laughter

Speech and laughter are uniquely human. Although there is considerable information on the neuronal representation of speech, little is known about brain mechanisms of laughter. Here we report that electrical stimulation in the anterior part of the human supplementary motor area (SMA) can elicit laughter. This area is also involved in the initiation of speech and has been shown to have increased activity in people who stutter¹.

Electrical stimulation was applied at 85 discrete sites on the cortical surface of the left frontal lobe of a 16-year-old girl (A.K.) undergoing monitoring by intracranial subdural electrodes to locate the focus of chronic intractable seizures. The patient's seizures were never accompanied by laughter. During stimulation A.K. performed a variety of tasks including naming of objects, reading a paragraph of text, counting, rapid alternating supination and pronation of the forearms, finger-to-thumb apposition, and alternating flexion and dorsiflexion of the feet.

A small area measuring about 2 cm × 2 cm was identified on the left superior frontal gyrus where stimulation consistently produced laughter (Fig. 1, red circles). The laughter was accompanied by a sensation of merriment or mirth. Although it was evoked by stimulation on several trials, a different explanation for it was offered by the patient each time, attributing the laughter to whatever external stimulus was present. Thus, laughter was attributed to the particular object seen during naming ("the horse is funny"), to the particular content of a paragraph during reading, or to persons present in the room while the patient performed a finger apposition task ("you guys are just so funny... standing around").

The duration and intensity of laughter increased with the level of stimulation current. At low currents only a smile was present, while at higher currents a robust contagious laughter was induced. It was also accompanied by the cessation of all activities involving speech or hand movements. At neighbouring sites, speech arrest was provoked without the evocation of laughter (Fig. 1, yellow circles), and at medially adjacent sites stimulation arrested all manual and speech activities (Fig. 1, blue circles). These sites were anterior to the area where electrical

stimulation evoked complex responses involving lower and upper extremities typical of the supplementary motor area (Fig. 1, black circles)^{2,3}. The area encompassing sites of laughter, speech arrest and manual activity extended to the dorsal convexity of the hemisphere, in agreement with current notions about the lateral border of the SMA⁴.

Pathological laughter has been described in several clinical conditions, including pseudobulbar palsy and gelastic seizures associated with lesions in the hypothalamus or temporal lobe^{5,6}. However, laughter in these conditions is rarely associated with the appropriate emotional experience, and often the condition of the patient (for example, a seizure) precludes reporting of the emotional experience. There are a very few reports of laughter evoked by electrical stimulation at cortical sites including the anterior cingulate and orbitofrontal cortex⁷ and the basal temporal lobe⁵. Arroyo *et al.* postulated dissociation of the motor program of laughter and the experience of merriment, localizing the



Figure 1 Magnetic resonance imaging three-dimensional surface reconstruction of left hemisphere of patient A.K., depicting sites where electrical stimulation evoked behavioural responses. Constant current electrical stimulation (1–12 mA, 0.3 ms duration, biphasic rectangular pulses), below the threshold of after-discharge, was applied at 50 Hz in a bipolar fashion for a duration of 5 s through adjacent contacts 1 cm apart on a subdural grid and strip electrodes implanted over the cortical surface of the left frontal lobe. No stimulation was performed at the right frontal lobe. Electrodes were implanted to identify the seizure focus, and electrical stimulation mapping was performed to plan the surgical resection of the focus. Informed consent was obtained in accordance with protocol approved by an internal review board. Key to colours: red, laughter; yellow, disruption or arrest of speech; blue, disruption or arrest of speech, naming and manual activity; black, motor movements involving the lower and upper extremities; green, tingling sensations in the right lower extremity. Reversal of the N20 potentials evoked by right posterior tibial nerve stimulation confirmed the location of the central sulcus at the anterior green circle.

former in the anterior cingulate and the latter in the temporal lobe⁵; however, our data show that laughter and mirth can be evoked by stimulation of the frontal cortex in the anterior part of the SMA, an area associated with the execution of motor programs.

The observation that A.K. was able each time to invoke a stimulus context that 'explained' the laughter suggests a close link between the motor, affective and cognitive components of laughter. It is likely that these varied components are represented in a large neuronal network capable of parallel distributed processing, where the entire network is activated as a whole by the stimulation of any of its constituent units⁸. Our observations also suggest that smiling and laughter might involve similar mechanisms and are closely related phenomena on a single continuum.

Although it could be argued that the neural representation of laughter observed in our epilepsy patient might not reflect the substrate for laughter in the normal brain, it should be emphasized that laughter was not a part of A.K.'s seizures. Furthermore, the region of seizure onset was medial to the area of evoked laughter, and stimulation in this region did not produce laughter responses. The human SMA has somatotopic organization, with the hands represented anterior to the arms and legs, and face and speech represented most anteriorly³. Our results suggest that speech and laughter are closely represented in the rostral part of the SMA, just anterior to the representation of manual activity.

We propose that the anterior part of the SMA is part of a further development in humans to accommodate the specialized functions of speech, manual dexterity and laughter. This area might correspond to the pre-supplementary motor area, a region situated anterior to the SMA proper, recently described in non-human primates, and thought to be involved in high-level motor programming^{9,10}.

Itzhak Fried, Charles L. Wilson, Katherine A. MacDonald, Eric J. Behnke
Division of Neurosurgery and Departments of Neurology and Psychiatry & Biobehavioral Sciences,
UCLA Medical School, Los Angeles, California 90095-7039, USA
e-mail: fried@surgey.medsch.ucla.edu

1. Fox, P. T. et al. *Nature* 382, 158–161 (1996).
2. Mitz, A. R. & Wise, S. P. *J. Neurosci.* 7, 1010–1021 (1987).
3. Fried, I. et al. *J. Neurosci.* 11, 3656–3666 (1991).
4. Wise, S. P. et al. *Adv. Neurol.* 70, 489–495 (1996).
5. Arroyo, S. et al. *Brain* 116, 757–780 (1993).
6. Martin, J. P. *Brain* 73, 453–464 (1950).
7. Sem-Jacobsen, C. W. in *Depth-electrographic Stimulation of the Human Brain and Behavior* (ed. Gant, W. H.) 127–138 (Charles C. Thomas, Springfield, Illinois, 1968).
8. Rumelhart, D. E. & McClelland, J. L. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* vol. 1 (MIT Press, Cambridge, Massachusetts, 1978).
9. Tanji, J. *Curr. Opin. Neurobiol.* 6, 782–787 (1996).
10. Picard, N. & Strick, P. *Cerebral Cortex* 6, 342–353 (1996).

Wobbly pursuit of extrasolar planets

Planet Quest: The Epic Discovery of Alien Solar Systems

by Ken Croswell

Oxford University Press/Free Press: 1997.

Pp. 324. £18.99, £25

Worlds Unnumbered: The Search for Extrasolar Planets

by Donald Goldsmith

University Science Books: 1997. Pp. 237.

\$28.50, £20.95. Distributed by W. H. Freeman

The Quest for Alien Planets: Exploring Worlds Outside the Solar System

by Paul Halpern

Plenum: 1997. Pp. 293. \$27.95, £16.94

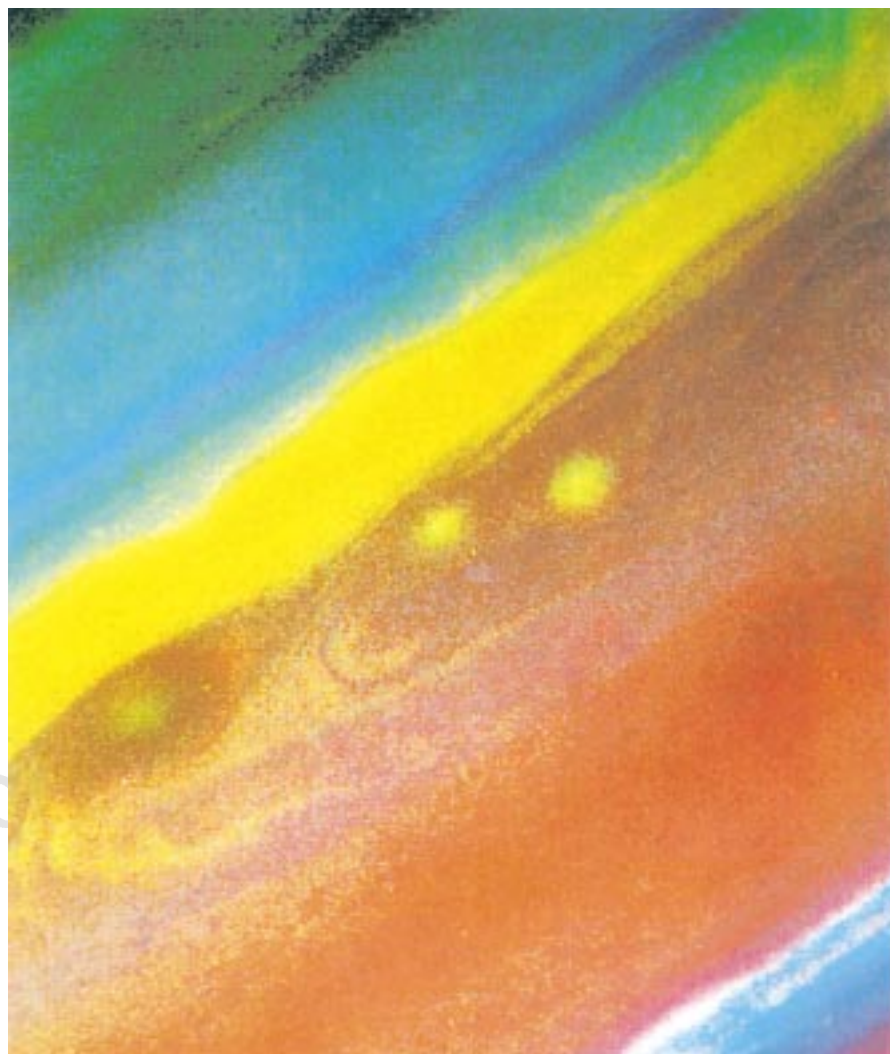
David W. Hughes

Planetary systems have been thought commonplace ever since Nicolaus Copernicus moved the Earth from the centre of the Universe, and Galileo Galilei converted the Sun into a normal star. I once suggested that about 4 per cent of all stars had planets. But where are they and what are they like?

Unfortunately in 1998 Earth-dwellers have instrumentation of such limited sensitivity that we have only vague hints about the characteristics of the few other planetary systems that have been found. And arguments still rage in some cases as to whether we have actually discovered planets or just found brown-dwarf stars.

There are seven ways of discovering extrasolar planets. Astrometry of close stars sometimes reveals a wobble caused by the gravitational influence of an unseen companion. Peter van de Kamp of the Sproul Observatory, Pennsylvania, took 2,000 photographs of Barnard's star between 1938 and the late 1960s and 'discovered' a wobble of 24.5 milliarcseconds. The suggested culprit was a planet 1.6 times heavier than Jupiter. But when Wulff Heintz and George Gatewood re-examined the star in the 1970s, it did not wobble. By 1996 Gatewood had shown that Lalande 21185 still did, and had two companions the size of Jupiter.

Planets can also be revealed by Doppler wobble which, unlike astrometry, is independent of distance. Stellar spectral lines oscillate as the star orbits the centre of mass of its planetary system. Jupiter, for example, causes the Sun's velocity to vary by about 12.5 metres per second over a 12-year period; astronomers can measure Doppler velocities to an accuracy of about 3 metres per second. Michel Mayor and Didier Queloz applied this technique to a host of Sun-like stars and in October 1995 discovered a planet around 51 Pegasi. This planet was half the mass of Jupiter, but was a mere 0.05, rather than the expected 5, astronomical units from the star's surface, and orbited every 4.2 days.



Planet spotting: three spots in Saturn's northern hemisphere, from Voyager 2 on 19 August 1981. The picture appears in the revised *Philip's Atlas of the Universe* by Patrick Moore. The previous edition was described in these pages as "Hugely informative... accessible to all interested readers". Philip's, £25.

Other Pegasian planets have been found around Rho Cancri A, Tau Boötis A, Upsilon Andromedae and Rho Coronae Borealis. The big shock is that these planetary systems are nothing like the one we inhabit.

Pulsar timing has also revealed orbiting planets; indeed, Alex Wolszczan found three planets (two Earth-like and one Moon-like) around the millisecond pulsar PST B1257+12. These systems are also unlike our own, and most astronomers think they were formed after the supernova phase.

More marginal discovery techniques rely on detecting planetary transits and on gravitational lensing. An observer in the jovian orbit plane would see Jupiter taking about 28 hours to transit the Sun every 12 years, and covering 1/96 of the Sun's surface area in the process. Gravitational microlensing occurs when two stars lie on the same line of sight. If the intervening star had a distant planet,

this would affect the microlensing light profile. Infrared interferometry could reveal planetary details but requires high-precision interferometers and is really a technique of the future. Finally we have the search for extraterrestrial intelligence. Messages from space would probably come from distant planet-dwellers. But we have not been contacted yet and will have to keep listening.

The 1990s is the decade of alien planetary systems. Pulsar planets were discovered (see *Nature* 352, 311; 1991) and then undiscovered (*Nature* 355, 213; 1992) and discovered again (*Nature* 355, 145; 1992). Doppler Pegasian planets were discovered (*Nature* 378, 355; 1995), confirmed by Geoffrey Marcy and Paul Butler and then found around other stars. This drama generated a huge media circus and spawned the three books under review, all of which have been aimed at the general reader and approach the

subject non-mathematically.

Of the three books, I like *Planet Quest* best. Ken Croswell has everything needed in a great science writer. He knows his subject and understands the joys, difficulties and frustrations of being an astronomer. He has great skill in interviewing the scientific protagonists and drawing from them the details of the excitement and disappointments of their endeavours; many of these interviews are quoted verbatim. And he can tell a superb story, which is compelling and complete. I particularly liked his discussion of the way the media reacted to the 'new planet' news. The book is riveting. And what is more, it has an extensive note and reference section.

Worlds Unnumbered comes a close second. Donald Goldsmith also discusses the new planets revealed in the 1990s but then places them in the wider context of the origin of life, the likelihood of extraterrestrial intelligence, and the perpetual fascination of humanity with unidentified flying objects. He also tries to predict where the subject will be at the end of the next two decades.

The *Quest for Alien Planets* by Paul Halpern is thinner than the other two and more prosaic. Halpern has clearly spent considerable time talking to people and picking up the cut and thrust between the observational astronomers, who are pushing their techniques to the limit, and the theoreticians, who have laboured long and hard to explain how our Solar System formed and

are greatly concerned that the new systems have very little in common with the one we live in and clearly had a radically different process of generation. He has produced an engaging tale. He stresses the possible role of planets as the elusive 'dark matter' and also looks to a very distant future when earthlings may have to find somewhere else to live.

It is clear from all these books that we still have a lot of searching to do. Planet hunters will not rest until they have in their sights some more small blue worlds that are warm and wet, and upon whose wind-whipped oceans and verdant forests shines a bright, constant, yellow star like our own. □

David W. Hughes is in the Department of Physics, University of Sheffield, Sheffield S3 7RH, UK.

Life laws

Physical Theory in Biology: Foundations and Explorations

edited by Charles J. Lumsden, Wendy A. Brandts and Lynn E. H. Trainor
World Scientific: 1997. Pp. 486. \$85, £59 (hbk); \$42, £21 (pbk)

PerBak

Is biology too difficult for biologists? And what can physics, dealing with the simple and lawful, contribute to biology, which deals with the complex and diverse?

It may not be useful to distil out general

laws of structure or behaviour from the phenomena studied by biologists. Nevertheless, in the words of the editors of this book, a marriage is being forged between the two sciences by the information revolution. In 20 essays, some of the foremost practitioners of the enterprise guide us through a cross-section of subjects in which the theory of physics might provide insight into biology.

The essays range from the sound and useful, through the contrived and irrelevant, to the speculative and imaginative. After a philosophical start in which the editors struggle with the meaning and definition of so-called fundamental concepts such as emergence, holism, complexity and computation (not generally important in our simple reductionist science of physics), the book gets up to speed with reviews of specific applications. It goes to show that we learn general principles from specific examples, not vice versa.

Various authors discuss pattern formation by introducing the 'morphogenetic field'. Patterns as diverse as the intricate surface features of the single-celled *Tetrahymena* and limb generation in the salamander can be described by minimizing an 'energy' function of an appropriately defined morphological field. The theory is rather phenomenological because the microscopic origin is largely unknown.

Goodwin then discusses two methods of deriving shape and form: minimizing an elastic energy function, and studying reaction-diffusion (forest fire) models. The first approach was initiated by D'Arcy Thompson as long ago as 1917, and the second by Alan Turing in 1952. The longevity of the processes is a direct measure of their success. Strikingly complicated forms can arise from very simple equations: the dream of theoretical physics come true! Another success story is the application of statistical mechanics to membranes and bilayers, which is (too) briefly covered by Jones and Tevlin. Here, concepts from theoretical physics are essential for any understanding.

The big targets for physics theorists are biological evolution and the brain. These complex many-body problems might have similarities to problems studied in particle- and solid-state physics. Busch and Trainor review Hopfield-type models of neuronal learning, but I see this as an example of physicists leading a whole field astray. It is hard to imagine a biological foundation for the complicated procedures for updating the synaptic strengths in those models. Nevertheless, the authors give a fair representation of the poor state of the art. They also apply their amusing generalization of the Hopfield neural network model to the organization of ants into eight different categories. This is an example of the way in which similar principles may apply in entirely different areas of science.

Some physical, or rather mathematical, considerations on topics in evolution are

Letters to Leibniz in Berlin

Ruth Tesmar, a German artist and professor of aesthetic practice, is celebrating the legendary letter-writing of the mathematician and rationalist Gottfried Wilhelm Leibniz in an exhibition that opens today at the Berlin-Brandenburg Academy of Sciences in Berlin. She explores the balance between aesthetics and scientific-technical rationalism. "Letters about monads", shown here, represents Leibniz's theory of monads by playing with writing, the traditional means of conveying theories. Tesmar's handwriting frames two images of Leibniz's work, indicating the separation of ideas by boundaries. Rice and a leaf symbolize paper-making materials. "Briefe an Leibniz" runs until the summer. Alison Abbott is the senior European correspondent of Nature.



then reviewed. Evolution on the basis of the statistics of DNA sequences is discussed, and Rowe explains the evolution of evolutionary stable strategies within the framework of game theory. The field of biological physics is vast, and this book barely touches a couple of corners.

The important issue here is not what physics theory has done for biology (which is not very much), but what it can do in the future, and to this end the book does a marvellous job of defining the arena. My own view is that biology is proving to be a goldmine for theoretical physicists at a time when large areas of traditional physics are almost exhausted. Read this book (and others) and get going! □

Per Bak is at the Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark.

Individual view

Metaphysics and the Origin of Species

by Michael T. Ghiselin

State University of New York Press: 1997.

Pp. 377. \$24.95, £19.50 (pbk).

Mark Ridley

Michael T. Ghiselin is an idiosyncratic mix of biologist, historian and philosopher. In 1969 he published *The Triumph of the Darwinian Method*, which is still, after 30 years of the Darwin industry, the book I should recommend to a scientific reader who wanted to understand Darwin's thinking. In the early 1980s, assisted by a MacArthur prize, he was one of the biologists who rescued molecular phylogenetics from its 20-year obsession with humans and chimpanzees and redirected it to transform evolutionary biology. But he is probably best known for his campaign, beginning with a paper in 1966, to establish that biological species are, in a philosophical sense, individuals.

It sounds an odd claim, but only because of a pun. As Ghiselin says, "in the usual biological sense, 'individual' is a synonym for 'organism' but the ontological term is a much, much broader one." He does not mean that biological species are organisms (which clearly they are not), but that they are logically — he would say metaphysically — individuals as opposed to classes. Classes are spatiotemporally unrestricted and have defining properties by which you can recognize members, or instances, of the class; individuals are spatiotemporally restricted, non-instantiable (you do not point and say "now there is a particularly fine example of Bill Clinton") and lack defining properties.

For instance, there is a class of all chairs (defined in some suitable way, such as all four-legged furniture that humans use to sit on), and there are individual members of the class. A particular chair has its own history: it

is built in a certain place, put in the library, has its arm broken by a student, is stained when a mug of coffee (existentially contrary to library rules) is spilled on it, and is finally destroyed by the gravity of a fat professor. Many scientific classifications do define classes: chemical elements are classes, for example gold consists of all elements with atomic weight 79. But are biological species like the class of chairs, or the particular chair? Most biologists before Ghiselin accepted that each species is a class: thus we speak of a person as being a 'member' of *Homo sapiens*, and species as having 'definitions' drawn up by museum-bench taxonomists.

Ghiselin noticed that, because of evolution, biological species cannot have defining properties. All swans may be white now, but if one mutated to black it would not cease to be a member of the swan species (assuming it could interbreed as normal). A species originates in a speciation event, evolves indefinitely down a lineage, and becomes extinct, analogously to that individual chair. The attributes that a species happens to possess are a contingent matter, like the attributes — the clothes or hairstyle — of a person. Individual attributes may help you to recognize someone at a certain time, but they do not logically define the person, because they can be changed without changing the person into someone else. Linnaean names, like *Homo sapiens*, are therefore proper names. A taxonomic description is logically a diagnosis not a definition.

In *Metaphysics and the Origin of Species* (the title deliberately alludes to the classics of Theodosius Dobzhansky and Ernst Mayr), Ghiselin makes his case fully, from first principles, and defends it against its many critics. In later chapters he explores the implications of the 'individuality thesis' in related topics, such as development and evolution, macroevolution and phylogenetics.

The book is of the kind that narrow-minded laboratory scientists call philosophical and like to ignore. "Why bother to think, when you can do experiments!" (*mutatis mutandis* as biology matures into a descriptive science). There are no experiments in Ghiselin's book, and none of his questions can be answered experimentally. But many scientists and philosophers do like to think about these things. What will they find here?

Critics will find it personal to the point of egocentric ("The possibility that species are processes, rather than wholes composed of organisms, has perhaps never occurred to anyone besides me"), discursive to the point of rambling ("Be that as it may..." is a common paragraph opener), rabbinical about historical texts, and polemical to the point of rude. Supporters will find it stimulating and enjoy the insights across an exceptional range of subjects, they will admire the erudition ("Curiously, the original 'Strickland Code' of zoological nomenclature, to which Darwin

New in paperback

The Invisible World: Early Modern Philosophy and the Invention of the Microscope

by Catherine Wilson

Princeton University Press, \$18.95, £14.95

"Wilson shows that microscopic observations reinforced the contemporary idea of the 'living machine' — that is, a reductionist view of nature. And therein lies the ultimate paradox of our machine-driven science: the essence of our natural world remains hidden despite our increasingly sophisticated scientific technology", Willem Hackmann, *Nature* 380, 35 (1996).

Subjected to Science: Human Experimentation in America Before the Second World War

by Susan E. Lederer

Johns Hopkins University Press, \$15.95, £13

"The author's description of experimentation on children and animals before the Second World War is much more complete than her analysis of research on prisoners and members of the military.... Here analysis of the legal cases, however, leaves much to be desired", George J. Annas, *Nature* 374, 827 (1995).

America's Pursuit of Precision Bombing, 1910–1945

by Stephen L. McFarland

Smithsonian Institution Press, \$19.95

McFarland traces the evolution of bombsights and automatic pilots from early efforts to the dropping of atomic ordnance on Japan in 1945.

What Will Be: How the New World of Information Will Change Our Lives

by Michael Dertouzos

HarperEdge, \$14

Speculations from the director of MIT's Laboratory for Computer Science.

Deeper: A Two-Year Odyssey in Cyberspace

by John Seabrook

Faber, £6.99

This literary internet travel book grew out of two articles written for *The New Yorker* magazine.

Fantasia Mathematica The Mathematical Magpie

edited by Clifton Fadiman

Copernicus, \$19.95 each

Two classic anthologies of mathematical stories, anecdotes, essays, rhymes, music, cartoons, aphorisms and other oddities, with authors ranging from Mark Twain and Lewis Carroll, to Aldous Huxley, George Gamow and H. G. Wells.

To the Arctic! The Story of Northern Exploration from Earliest Times

by Jeannette Mirsky

University of Chicago Press, \$16.95, £13.50

was a signatory, actually refers to the names of taxa as proper names”), and be excited by the vigour of the prose from which Ghiselin’s personality is inseparable. I waffled (in the oscillatory American sense) between the two. Either way, it is now the standard source for anyone who wants to find out what it means to call a species an individual. □

Mark Ridley is in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

Novel adaptation

Enduring Love

by Ian McEwan

Cape/Doubleday: 1997. Pp. 247. £15.99, \$23.95

Marek Kohn

If our minds contain modules adapted to detect cheating, as the evolutionary psychologist Leda Cosmides has proposed, we run the risk that these organs will reach uncomfortable conclusions about the individuals they have most opportunity to observe: ourselves. The people who are best at deluding others are those who are best at deluding themselves, reflects Joe Rose, the narrator of *Enduring Love*; they flourish, and so do their genes. But his powers of persuasion are not so well developed. His genes have not flourished either, because this is a Darwinian morality tale.

His commercial timing as a writer has been perfect, however. He has surfed every wave in the popular science tide: dinosaurs, black holes, quantum magic, superstrings, neuroscience, Darwin revisited.... But now it is too late for him to follow the career in scientific research that, according to the voice of his conscience, was his true calling. At one point he writes an article arguing that Victorian scientists were the last to enjoy the pleasure of narrative in their papers, and then realizes it does not stand up; it was not written in the pursuit of truth; it was not science but journalism. Ian McEwan also takes his science seriously, and for many years regretted doing English rather than a science degree.

Despite carrying too much guilt, too much bulk and too little hair, Joe enjoys the love of a beautiful young woman. He cannot quite believe his luck, but no neo-Darwinian mate-choice theorist would be surprised. And his partner Clarissa is unable to have children. The couple remain on a plateau of erotic arousal, enjoying endless proximal rewards without an ultimate goal, cheating the evolutionary system as well as the intellectual one.

Then, one day in the countryside outside London, Clarissa and Joe see a balloon in trouble. Joe is one of several men who throw themselves into the scene, assembling a fatally *ad hoc* male coalition. Any primatologist will tell you that if you were a boy about to be car-



The social life of baboons

The hamadryas baboon was held sacred by the ancient Egyptians, but its highly complex social organization makes it just as intriguing to modern primatologists such as Hans Kummer, whose fascinating book about the hamadryas, *In Quest of the Sacred Baboon: A Scientist's Journey* (Princeton, \$16.95, £12.95), is out now in paperback. “Truly multidimensional.... It is a distillation of decades of making sense of the most complex social organization of any species of non-human primate.... From the arresting first page of the preface, which takes the reader straight into the bush, to the moving final paragraph that will resonate with long-term field workers, it is compelling.... The thoughtful, personal disclosures are uncommonly telling, and deserve wide attention”, W. C. McGrew, *Nature* 379, 410 (1996).



ried away in a balloon, you would be misguided to place your trust in a coalition of unrelated males who did not know each other. As the balloon rises, one man lets go. The balance of interests is upset: the value of cooperation collapses, and that of defection soars. Others then drop off, including Joe, but the last man hangs on. He is killed; the boy survives.

As well as the death, there is a second cataclysmic moment within this drama. One of the rescuers is a young man, Jed Parry, whose unconditional religious faith is suddenly joined by an equally absolute conviction that he and Joe are destined to be together in Christ. Joe diagnoses de Clérambault’s syndrome, in which the subject believes the object of the obsession to be in love with him, however strong the evidence to the contrary. Jed’s menacing ecstasy batters away at Joe’s sanity, and pulls apart the bonds between Joe and Clarissa. On one side is the narrator’s voice of reason; on the other are the Jesus freak’s deranged faith and the Romantic inclinations of Clarissa, a student of Keats. Science finds both art and religion ranged against it.

As narrator, however, Joe has also been given the benefit of comprehensive emotional literacy, in reflection if not in dialogue. He fails to realize, for instance, that barren Clarissa might not care to hear his speculations about the adaptive value of a baby’s smile. Joe is particularly taken with Darwin’s *Expression of the Emotions* and Paul Ekman’s recent work in the same field.

These scientific interpolations work better than the obvious oppositions between science and unreason. But the most effective science in *Enduring Love* is that which is not completely spelled out. The dynamics of the balloon coalition are one instance; another is the inexorable intrusion of reproductive interests, subtly evoked in Joe’s encounters with the children of the dead man.

McEwan’s final scientific flourish is an appendix, a cod scientific paper that recapitulates the story as a case history of de Clérambault’s syndrome. Into this he drops a throwaway happy ending, thereby nailing the lid on Joe’s vain thesis about science losing the plot. □

Marek Kohn is at Flat 3, 27 Warrington Crescent, London W9 1ED, UK.

The motion of crustal blocks driven by flow of the lower lithosphere and implications for slip rates of continental strike-slip faults

S. J. Bourne, P. C. England & B. Parsons

Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK

Geodetic measurements in actively deforming areas of the continents reveal the pattern of deformation in the lithosphere. If the dominant forces acting on crustal blocks are tractions at their bases, then the long-term motion of each block will be given by the average velocity of the underlying lithosphere. Slip rates between blocks estimated in this way from recent geodetic measurements across fault zones in the South Island of New Zealand and Southern California are in good agreement with slip rates estimated geologically.

Unlike oceanic plates, which move rigidly across the surface of the Earth, continental lithosphere is weak in comparison with the forces driving plate motion, and deforms over wide regions at plate boundaries. The lower crust and upper mantle of the continental lithosphere probably deform in a ductile fashion^{1,2}, but the upper crust, on which observations of active deformation are made, is discontinuous, with slip on faults accommodating the relative motions of crustal blocks.

Studies of the depth extent of rupture in large continental earthquakes^{3,4} show that seismic slip on these faults occurs mainly in the upper 10–20 km of the crust. Opinions are divided as to how deformation is accommodated below this level. In one view, the faults cut through the entire mechanically strong part of the continents, as they do through oceanic lithosphere. Deformation occurs, in the upper layer, by stick–slip motion on faults, and, in the lower layer, by stable sliding on narrow shear zones beneath the seismogenic surface faults^{5–7}. In an alternative view, the lower part of the lithosphere deforms in a distributed fashion, applying shear tractions to the base of the brittle upper-crustal layer. The discontinuous deformation at the surface reflects localization, onto discrete fault surfaces, of deformation that is distributed at depth^{2,8–12}.

Here we investigate some consequences of the second viewpoint. If the tractions applied to the bases of crustal blocks, by the ductile layer beneath them, exceed in magnitude the shear tractions acting on the faults bounding the blocks, then the blocks will passively follow the pattern of deformation in the underlying continental lithosphere^{9,12}. It follows that, if the pattern of deformation in the lower lithosphere can be determined from geodetic measurements, the motion of the crustal blocks, and hence the relative motion between them, can be predicted. We test this idea for two plate-boundary zones where distributed crustal deformation occurs in a relatively simple configuration: the Marlborough fault zone in the South Island of New Zealand, and Southern California, south of the Transverse Ranges. We find that the slip rates on the faults predicted from the geodetic measurements are in good agreement with geologically determined slip rates.

The basic model

Each of the regions we consider is deforming by distributed right-lateral shear, with the long-term surface motion taken up on a number of strike-slip faults orientated approximately parallel to the relative plate motion across the zone. For this reason, we restrict our analysis to the simple geometry displayed by these zones, and adopt

the idealization that horizontal displacements and velocities do not vary in the direction parallel to the faults (Fig. 1). We assume that the upper crust is coupled to its ductile substrate; therefore, if transient effects due to earthquake activity can be ignored, the velocity of upper crust will, over the time interval between earthquakes, be the same as the long-term velocity of the lithosphere below it, although its strain will be elastic, rather than permanent (Fig. 1a). But because earthquakes repeatedly release the elastic strain, the crustal blocks move (over intervals much longer than the time between earthquakes) as unstrained bodies, each with its own velocity. Differences in velocity between adjacent blocks represent long-term slip rates on the faults separating the blocks. Over the

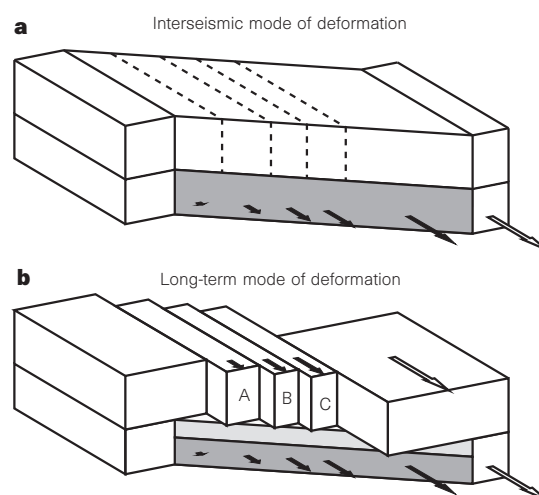


Figure 1 Interseismic and long-term kinematics of crustal blocks within a zone of simple shear. **a**, Between earthquakes, the faults remain locked and the crustal blocks accumulate elastic strain through the action of basal stresses applied by the viscous layer beneath. (The elastic strain is greatly exaggerated here; shear strains in the interseismic period are usually $\sim 10^{-5}$.) **b**, The finite strain over many seismic cycles consists of ductile shear in the lower lithosphere, and block-like motion of the upper crust. The long-term velocity of each block is the same as the average velocity of the fluid beneath it. The transition between block motion at the surface and distributed motion at depth takes place over a small depth interval, shown in lighter shading. Details of this transition are shown in Fig. 2.

long term, therefore, the velocity at a point on a block will in general differ from that in the ductile layer beneath it (Fig. 1b).

A simple argument suggests that the long-term velocity of a crustal block should equal the average velocity of the ductile lithosphere beneath it. If the frictional forces on faults either side of a block are negligible in comparison with basal tractions, the block will be in equilibrium when the net drag force on its base is zero. The basal drag force arises from the vertical gradient of the velocity beneath the blocks^{12,13} (Fig. 2). The drag is, therefore, proportional to the difference in velocity between a point on the base of the block and the velocity of the smoothly deforming lithosphere at depth beneath it (Fig. 2). The condition that the net drag force on the base of the block be zero is met if this velocity difference integrated over the whole block is zero, that is, if the long-term velocity of the block is the same as the average velocity of its underlying ductile lithosphere. The difference in velocity of adjacent blocks then gives the slip rate on the fault between them.

The time-averaged motion beneath the blocks must vary with depth, from that of the blocks to that of the smooth deformation at depth (Fig. 1). The thickness of the zone over which this transition occurs can again be estimated from a simple argument. Consider a case in which the velocity profile at depth in the ductile lithosphere is linear (constant horizontal gradient of velocity), and the surface is divided into a number of blocks of equal width (Fig. 2). If the fluid below the blocks is newtonian, then the variation of the velocity u parallel to the faults is described by Laplace's equation, $\partial^2 u / \partial x^2 + \partial^2 u / \partial z^2 = 0$, where z is depth and x is distance perpendicular to the faults. The difference between the step-like velocity profile of the crustal blocks and the linear gradient of the substrate is a periodic saw-toothed function, of wavelength a , the width of the crustal blocks (Fig. 2). The solution to Laplace's equation subject to

these conditions can be obtained using Fourier series, and it can be shown that the amplitude of the velocity difference decays over a depth scale that is $\sim a/\pi$ (Fig. 2). For block widths of 15–20 km (Figs 3 and 4), the transition occurs over a depth interval of 5–7 km, justifying the depiction, in Fig. 1b, of a thin transition zone between the long-term displacements of the crustal blocks and the smooth displacements of the lower lithosphere. If the ductile lithosphere has a power-law rheology, thought to be characteristic of the creep of rocks at lithospheric stresses¹, shear would be concentrated even more strongly just below the base of the blocks than in the case of a newtonian fluid.

Given the proposition that the crustal blocks are driven by tractions on their bases, then geodetic measurements of surface velocity will represent the long-term flow of the deeper lithosphere, provided that the surface motions have not been disturbed by the transient effects of a recent earthquake—a condition we discuss further below. Because, in the long term, the time-averaged velocity of a crustal block will be equal to the velocity of the deeper lithosphere beneath it, the surface measurements made over the short term should yield the long-term average velocities of the crustal blocks, and hence the rates of slip on the faults that separate them. To test this proposition we have compared slip rates predicted using recent geodetic measurements in the South Island of New Zealand and Southern California with geological estimates of slip rates on the faults in these areas.

Geodetic measurements of lithospheric deformation

Reliable methods for the geodetic determination of crustal deformation are well established and have been widely applied to regions of active continental deformation. The geodetic data considered here comprise measurements using the Global Positioning System (GPS) and terrestrially based methods of triangulation and trilateration.

Within the Marlborough region of the Southern Island of New Zealand, a triangulation and trilateration network was observed in

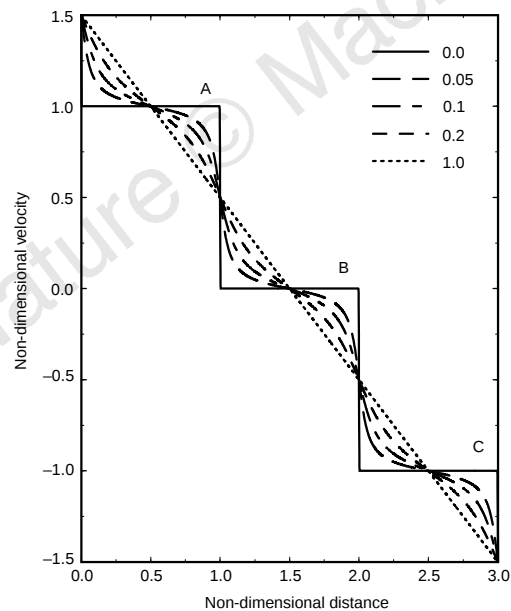


Figure 2 Variation with depth of velocity parallel to the faults. Velocities have been made non-dimensional by dividing them by the velocity difference between blocks, taken to be constant in this example, and are measured relative to the central block. The blocks impose a time-averaged velocity profile on the top of the ductile layer that is step-like (solid line). The figure illustrates the velocity profile that would be generated in a newtonian fluid subjected to this velocity condition on its upper surface, with a linear velocity gradient at depth. Distances have been made non-dimensional by dividing them by the width of the blocks, and the velocity is shown at a set of non-dimensional depths, 0.0–1.0, measured from the base of the blocks. The labels A, B, and C correspond to those in Fig. 1. The time-averaged strain rate is effectively uniform for non-dimensional depths greater than ~ 0.3 .

Table 1 Calculated and measured fault slip rates

Fault	Slip rates (mm yr ⁻¹)		Ref.
	Calculated	Measured	
Northern South Island, New Zealand			
Hope	19	20–25	38
		11–17	34
		19 ± 5	36
Clarence	5	1–8	37
		4–8	38
Awatere	6	5–7	37
		5–10	38
		6–7	39
Wairau	9	4 ± 1	35
		3.7–10.7	37
Southern California, United States			
San Andreas	22	25 ± 5	31
San Jacinto	12	12 ± 7	32
Elsinore	6	5 ± 2	32
Palos Verdes and Newport-Inglewood	6	5–6	32, 49, 50
San Clemente	3	1–4	33

'Calculated' slip rates show comparisons between slip rates calculated from geodetic measurements of crustal velocity, under the assumptions described in the text. Slip-rate measurements in the South Island of New Zealand are determined from the age of offsets based on either the time of the Last Glacial Maximum^{34,35} or Knuepfer's weathering-rind calibration curve^{36,37}. Slip rates reported by Knuepfer³⁷ all show marked increases with age. His data for the Awatere fault are consistent with a constant slip rate of 5–7 mm yr⁻¹ although he published a temporally variable rate of 3.6–7.3 mm yr⁻¹; his data for the Hope fault show a much greater range in rate of 1.3–42.9 mm yr⁻¹. Slip rate measurements on the Elsinore, San Jacinto, and San Andreas faults are made from palaeoseismic data alone^{31,32}. The slip rate assigned to the combination of the Newport-Inglewood and Palos Verdes faults comes from ~ 3 mm yr⁻¹ estimated from offset marine features⁵⁰ and uplifted marine terraces⁴⁹, and from the presumption³² that 5–6 mm yr⁻¹ of slip on the Agua Blanca fault and the Coronado and Descanso-Rose Canyon faults directly to the south of the region is distributed between these two faults.

1982 under the Earth Deformation Programme of the Royal Society of New Zealand¹⁴. This narrow northwest–southeast orientated network crosses the Wairau, Awatere, Clarence and Hope faults of the Marlborough fault zone between the Pacific and Australian plates in a direction approximately perpendicular to their strike. Network solutions obtained from a two-dimensional adjustment of the 190 horizontal angles and 122 geodolite distance observations for 28 geodetic stations produce normally distributed residuals of 4.4×10^{-6} rad (about 1 arcsec) in horizontal angle and 64 mm in distance at the 1σ confidence level. These uncertainties should be compared with relative displacements of about 400 mm across the network (Fig. 3) over the interval 1982–94.

During a 1994 (days of year 024–041) GPS experiment¹⁵, 21 sites within this original network were re-occupied. Uncertainties in the

1994 coordinates can be estimated from the residuals between individual daily network solutions for the coordinates and the final network solution; these are normally distributed with 1σ confidence intervals of 5.6 mm in the northward, and 7.3 mm in the eastward, components of position.

The displacement of each site over the 12-year interval is calculated as the difference between its 1994 coordinates and the coordinates determined from the 1982 survey. Obtaining these latter coordinates entails making some assumptions. GPS measurements provide sufficient information to obtain relative position vectors for each site, whereas the terrestrial measurements provide information only about the horizontal angles and distances between the sites. We provide position and orientation for the 1982 terrestrial network by assuming that its most southeasterly site has remained fixed with respect to the Pacific plate, and that a baseline between two of the most northwesterly sites has not rotated with respect to the Australian plate. We also solved for a possible scale discrepancy between GPS and geodolite measurements by assuming that the most northwesterly baseline, where the velocity profile suggests strain rates are small, has not changed in length. The level of uncertainty in the horizontal velocity estimates increases monotonically with distance away from the fixed site; the average 1σ uncertainty in velocity is 6 mm yr^{-1} . Similar displacements were obtained when the orientation of other lines at the northwestern end of the network were fixed; the scale correction applied is smaller than the *a priori* uncertainty assigned to individual geodolite measurements.

A comprehensive set of geodetic data covering Southern California has been compiled from geodolite^{16–19} and GPS^{20–22} measurements over the period 1970–95, and combined to estimate horizontal site velocities for the field of contemporary interseismic deformation, by the Crustal Deformation Working Group of the Southern California Earthquake Center^{23,24}. The data are adjusted jointly to produce a single horizontal velocity estimate for each site under the constraint that velocities at any sites separated by less than 2 km are equal: 1σ uncertainties in the horizontal component of velocity are everywhere less than 5 mm yr^{-1} .

We consider here only those sites in the region south of the Transverse Ranges and north of the southern terminus of the San Andreas fault, so that the geometry of the crustal blocks in the region reflects the simple geometry indicated in Fig. 1; the choice also excludes regions affected by co- and post-seismic motion of the 1992 Landers earthquake. This set of 71 sites spans the sub-parallel San Clemente, Newport-Inglewood, Palos Verdes, Elsinore, San Jacinto, and San Andreas faults in a region where there is little along-strike change in the style of faulting.

Comparison between estimates of fault slip rates

Geodetic estimates of slip rate. The geological observations show that, in each region, a single fault accommodates nearly half the relative plate motion (the Hope and San Andreas faults), while the other faults are less active (Table 1). This fact is not immediately obvious from the geodetic observations, which show a smooth variation of velocity across each region (Figs 3 and 4). However, fault slip rates can be calculated from these velocity profiles, using the framework described in Fig. 1. On the assumptions discussed above, we calculate the time-averaged velocity of each crustal block by averaging the geodetically observed velocity across the block. To ensure that the distribution of data points does not bias the calculation, average velocities are calculated from smooth cubic splines fitted to the data by least-squares. Slip rates on the faults are calculated from the differences in velocity of adjacent blocks (Fig. 1).

An important qualitative result is that this calculation, which is completely independent of the geological estimates of slip rates, predicts the slip rates to be much higher on the Hope and San Andreas faults than on the other faults in the respective regions (Table 1). This result follows directly from the geometry of the faults

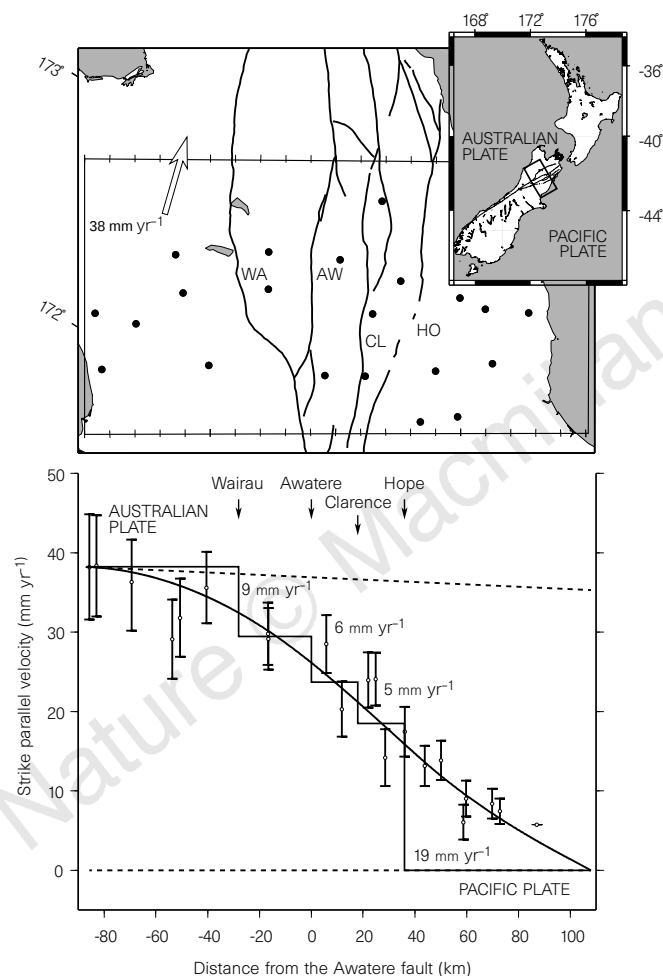


Figure 3 Observed crustal displacements between 1982 and 1994 in the Marlborough fault zone in the South Island of New Zealand, expressed as average velocities over the interval. The region of interest is denoted by the box shown both in the inset and in the main map (top). The main map is orientated according to the mean strike of the faults (060°) and distances normal to this direction are indicated by tick marks placed every 10 km. The component of surface velocity, parallel to 060° , observed at each site is plotted in the bottom panel as a function of distance from the Awatere fault; error estimates are shown at the 1σ level. Relative plate motion according to NUVEL-1A^{51,52} is indicated by the white arrow on the map, and its magnitude is denoted by dashed lines on the velocity profile. The long-term, time-averaged velocity profile across the fault blocks (thick solid line) is calculated as the average of the smoothed interseismic velocity across each block (thin solid line). Labels by each riser in the stepped velocity profile indicate the slip rates on the faults bounding the blocks, calculated from the smoothed velocity profile. Labels WA, AW, CL and HO refer, respectively, to the Wairau, Awatere, Clarence and Hope faults.

and blocks themselves. For example, the region east of the Hope fault is assumed to be attached to the Pacific plate. The mid-point of the block between the Hope and Clarence faults lies about 70 km from the eastern end of the region where there are significant velocity gradients (Fig. 3). In contrast, the mid-point of the block bounded by the Hope and Clarence faults and that bounded by the Clarence and Awatere faults are separated by about 20 km. Given that the strain rate across these faults is approximately uniform, when accumulated strain is released in earthquakes, the relative displacement across the Clarence fault will be smaller than that across the Hope fault because of the smaller distance over which the strain accumulated. A similar argument applies to Southern California.

These estimates of fault slip rate could be in error if the geodetic measurements of strain rate are contaminated by transient deformation following large earthquakes. In California, at least, post-seismic strain rate concentrations close to faults appear to decay to long-term average strain rates on a timescale of ~20–40 years (refs 7, 25). In the area of Southern California studied, there have been, over the past 100 years, one large earthquake (moment magnitude $M_w = 6.2$, in 1933) on the Newport-Inglewood fault, and eight on

or near the San Jacinto fault²⁶. Of the latter, the most recent large earthquakes, both $M_w = 6.5$, were in 1968 and 1987. Over the same period there have been two large earthquakes near the San Andreas fault in Southern California, $M_w = 6.0$ in 1948 and $M_w = 6.2$ in 1986²⁶. The largest earthquakes in or adjacent to the Marlborough fault zone were the 1888 North Canterbury earthquake^{27–29} (surface wave magnitude $M_s = 7.3$) on the Hope fault, and the $M_s = 7.8$, 1929 Murchison earthquake²⁹. The most recent large earthquakes near or within the network across the Marlborough faults were the $M_s = 6.4$, 1960 event in Southern Nelson²⁹ and the $M_w = 5.9$ Lake Tennyson earthquake in 1990³⁰. Coseismic movement associated with the latter event, which occurred within the interval of measurement, may have resulted in a small disturbance from the trend in Fig. 3. Otherwise there have been few large earthquakes in the areas considered within the past 30 years, and the observations show no concentrations in the strain rate distributions that might be due to earthquake-related strains (Figs 3 and 4).

Geological estimates of slip rate. Geological estimates of time-averaged slip rates on faults are available for the dating of measured offsets in features that cross the faults in the Marlborough fault zone and in Southern California. These offsets are divided by ages assigned to the offset features to yield slip rates.

The most reliable estimates are derived from the precise dating of small offsets using ^{14}C . Such estimates have been made for the San Andreas, the San Jacinto, and the Elsinore faults^{31,32}. In each case the rate calculated from the geodetic velocities agree within uncertainty, and to better than 3 mm yr^{-1} , with the geological estimates of slip rate (Table 1). Elsewhere in Southern California the slip rates are more uncertain. The combined slip on the Palos Verdes and Newport-Inglewood faults, which lie very close to each other, is estimated³² to be in the range $5\text{--}6 \text{ mm yr}^{-1}$, which agrees with the calculated slip rate for a single fault at this location (Table 1). The slip rate on the San Clemente fault is unknown, although Ward and Valensise³³ suggest a rate of $1\text{--}4 \text{ mm yr}^{-1}$, in agreement with the rate we calculate.

In New Zealand, ages have been assigned to surfaces containing offsets on the scale of 10–100 m either by dating them to the time of the Last Glacial Maximum, $17.2 \pm 2 \text{ kyr ago}$ ^{34,35}, or by measuring the thicknesses of weathering rinds developed by boulders exposed at the surface^{36,37}. Doubt must exist over the reliability of the calibration curve employed in the latter method, because all the slip rates calculated using this technique exhibit marked increases with age (Table 1), whereas no other evidence substantiates such large temporal changes in slip rates.

The geological estimates of slip rate on the Hope fault are mostly close to 20 mm yr^{-1} , although some estimates made from weathering rinds range substantially lower than this (Table 1). The estimate of slip rate calculated from the geodetic observations is 19 mm yr^{-1} . The geologically estimated rate on the Clarence fault of $4\text{--}8 \text{ mm yr}^{-1}$ published by Van Dissen and Yeats³⁸ is in agreement with the rate of 6 mm yr^{-1} calculated from the geodetic data; the range of $1\text{--}8 \text{ mm yr}^{-1}$ based on weathering rids is also consistent, but not definitively so. The estimates of slip rate on the Awatere fault from Van Dissen and Yeats³⁸ are $5\text{--}10 \text{ mm yr}^{-1}$, and the dates estimated using weathering rinds³⁷ give time-independent slip rates of $5\text{--}7 \text{ mm yr}^{-1}$ and time-dependent slip rates of $3.6\text{--}7.3 \text{ mm yr}^{-1}$. These estimates, and the more recent estimates of Little *et al.*³⁹ of $6\text{--}7 \text{ mm yr}^{-1}$, agree with the rate of 6 mm yr^{-1} calculated from the geodetic velocities. The slip rate calculated for the Wairau fault from the geodetic data is consistent with the upper limit of the weathering-rind estimate³⁷ but not with that based on assigning a date of 17 kyr ago to the offset features³⁵.

Implications for continental deformation

Continental tectonics bears a superficial resemblance to plate tectonics, in that the deformation, in some places at least, appears to be concentrated on a few major faults. It is tempting, therefore, to

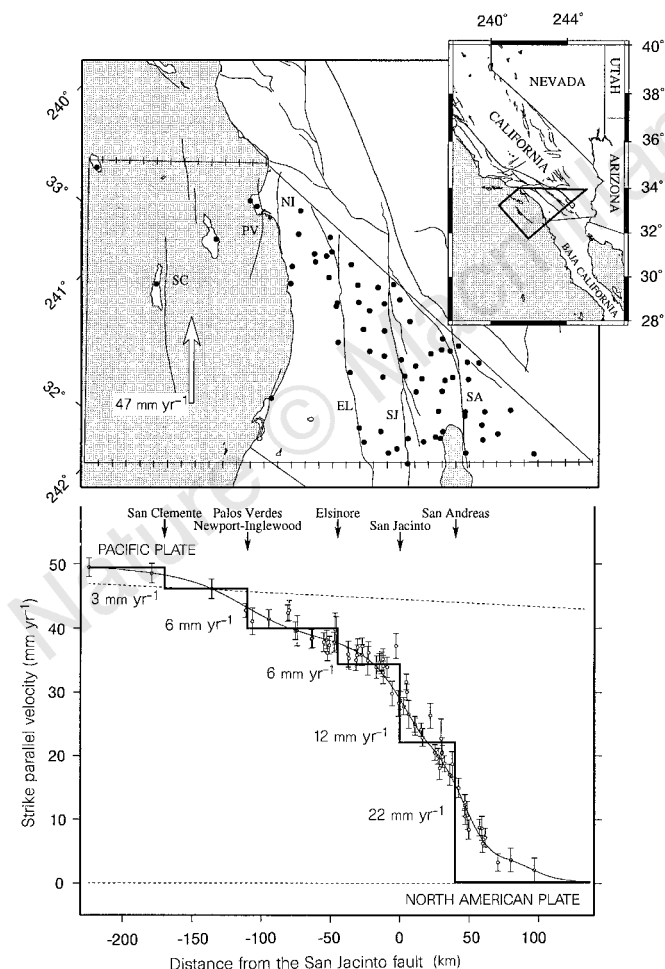


Figure 4 Geodetically observed crustal displacements in the interval 1970–95 in Southern California, south of the Transverse Ranges, shown in the same manner as Fig. 3. The plate relative motion is according to NUVEL-1A^{51,52} and the mean strike of the faults is taken as -041° . The velocities plot relative to three Very Long Baseline Interferometry stations on the North American plate. The trapezoidal shape of the region of study is chosen to exclude the Western Transverse Ranges, where thrust faulting and block rotation complicate the velocity field⁶³. Labels SC, PV, NI, EL, SJ and SA refer, respectively, to the San Clemente, Palos Verdes, Newport-Inglewood, Elsinore, San Jacinto and San Andreas faults.

describe the kinematics of the deformation in discontinuous terms. There are, however, two important qualitative differences between oceanic plates and continental crustal blocks. First, motions of adjacent continental fault blocks are often highly correlated (see, for example, Figs 3 and 4), whereas the motions of adjacent plates are independent of one another. Second, the widths of continental blocks are often small compared with the thickness of the lithosphere (~100 km), whereas even the smallest plates are many times wider than they are thick. These observations have led many authors to suggest that the organized motion of continental crustal blocks is the result of shear tractions applied to their bases by ductilely deforming lower lithosphere that is at least as strong as the discontinuous upper layer^{9,10,12,40–42}.

In the model commonly used to interpret geodetic observations of strain in fault zones, however, the surface deformation reflects elastic strain accumulation produced by continuous slip on the extensions at depth of faults that are locked near the surface^{7,43,44}. From this viewpoint, the slip rate on each fault must be specified before the resulting deformation can be calculated. In contrast, a consequence of the point of view adopted here is that the motions of the crustal blocks, and hence the slip rates on faults separating them, are determined by the velocity field in the lower lithosphere. In the absence of perturbation due to recent earthquakes, the short-term motion of the surface reflects the motion of the deeper lithosphere, and so may be used to calculate fault slip rates. We have tested this argument by comparing the observed long-term slip rates on major faults with those calculated from the distributions of faults and of geodetically determined surface velocities. We find that the fault slip rates predicted in this way for the Marlborough fault zone in the South Island of New Zealand and the fault zone in Southern California are consistent with estimates made from geological observations.

These results also bear on the accumulation and release of strain within a seismic cycle. Savage⁴⁵ pointed out that geodetic measurements alone cannot distinguish between distributed ductile flow and a time- and depth-dependent slip history on a single planar fault as explanations for the effects of earthquakes on crustal deformation. But the agreement between observed fault slip rates and those calculated here strongly supports the view that flow in the ductile portion of the lithosphere drives the accumulation of strain in the brittle upper layer, where faulting occurs^{7,8,41,42,46}, and must play a role in post-seismic relaxation of strain resulting from slip during earthquakes. The agreement also casts doubt on the suggestion^{47,48} that the strength of faults in California controls the orientation of the stress field there. Like Molnar², we suspect that stress orientations in the shallow crust merely reflect local adjustments to the kinematic requirement that the upper-crustal blocks must follow the flow of the lower layer. □

Received 17 June; accepted 11 November 1997.

1. Brace, W. & Kohlstedt, D. Limits on lithostatic stress imposed by laboratory experiments. *J. Geophys. Res.* **85**, 6248–6252 (1980).
2. Molnar, P. in *Fault Mechanics and the Transport Properties of Rocks: a Festschrift in Honor of W. F. Brace* (eds Evans, B. & Wong, T.-F.) 435–460 (Academic, London, 1991).
3. Chen, W. & Molnar, P. Focal depths of intercontinental and interplate earthquakes and their implications for the thermal and mechanical properties of the lithosphere. *J. Geophys. Res.* **88**, 4183–4214 (1983).
4. Jackson, J. & White, N. Normal faulting in the upper continental crust: observations from regions of active extension. *J. Struct. Geol.* **11**, 15–36 (1989).
5. Savage, J. & Burford, R. Accumulation of tectonic strain in California. *Bull. Seismol. Soc. Am.* **60**, 1877–1896 (1970).
6. Thatcher, W. Systematic inversion of geodetic data in central California. *J. Geophys. Res.* **84**, 2283–2295 (1979).
7. Thatcher, W. Nonlinear strain buildup and the earthquake cycle on the San Andreas fault. *J. Geophys. Res.* **88**, 5893–5902 (1983).
8. Prescott, W. H. & Nur, A. The accommodation of relative motion at depth on the San Andreas fault system in California. *J. Geophys. Res.* **86**, 999–1004 (1981).
9. McKenzie, D. & Jackson, J. The relationship between strain rates, crustal thickening, palaeomagnetism, finite strain and fault movements within a deforming zone. *Earth Planet. Sci. Lett.* **65**, 182–202 (1983).
10. Houseman, G. & England, P. Finite strain calculations of continental deformation. 1 Method and general results for convergent zones. *J. Geophys. Res.* **91**, 3651–3663 (1986).
11. England, P. & Molnar, P. Inferences of deviatoric stress in actively deforming belts from simple physical models. *Phil. Trans. R. Soc. Lond.* **337**, 151–164 (1991).

12. Lamb, S. Behaviour of the brittle crust in wide plate boundary zones. *J. Geophys. Res.* **99**, 4457–4483 (1994).
13. Lachenbruch, A. & Sass, J. Heat flow from Cajon Pass, fault strength, and tectonic implications. *J. Geophys. Res.* **97**, 4995–5015 (1993).
14. Bevin, A. *Misc. Series* **5**, 87–96 (1981).
15. Bourne, S. *et al.* Crustal deformation of the Marlborough Fault Zone in the South Island of New Zealand: Geodetic constraints over the interval 1982–1994. *J. Geophys. Res.* (in the press).
16. Prescott, W., Savage, J. & Kinoshita, W. Strain accumulation rates in the Western United States between 1970 and 1978. *J. Geophys. Res.* **84**, 5423–5435 (1979).
17. Savage, J., Prescott, W., Lisowski, M. & King, N. Deformation across the Salton Trough, California, 1973–77. *J. Geophys. Res.* **84**, 3069–3079 (1979).
18. Savage, J., Prescott, W. & Gu, G. Strain accumulation in southern California, 1973–1984. *J. Geophys. Res.* **91**, 7455–7473 (1986).
19. Lisowski, M., Savage, J. & Prescott, W. The velocity field along the San Andreas fault in central and southern California. *J. Geophys. Res.* **96**, 8369–8389 (1991).
20. Larsen, S. & Reilinger, R. Global Positioning System measurements of strain accumulation across the Imperial Valley, California: 1986–1989. *J. Geophys. Res.* **97**, 8865–8876 (1992).
21. Dong, D. *The Horizontal Velocity Field in Southern California from a Combination of Terrestrial and Space-geodetic data*. Thesis, Massachusetts Inst. Technol. (1993).
22. Feigl, K. *et al.* Space geodetic measurement of crustal deformation in central and southern California, 1984–1992. *J. Geophys. Res.* **98**, 21677–21712 (1993).
23. Shen, Z.-k. *et al.* *Horizontal Crustal Deformation Velocity Map Version 1* (Crustal Deformation Working Group, Southern California Earthquake Center); available at <http://minotaur.ess.ucla.edu/velmap/welcome.shtml>
24. Shen, Z. *et al.* Crustal deformation measured in Southern California. *Eos* **78**, 477 (1997).
25. Scholz, C. *The Mechanics of Earthquakes and Faulting* 1st edn (Cambridge Univ. Press, 1990).
26. Ellsworth, W. in *The San Andreas Fault System, California* (ed. Wallace, R.) 153–185 (Prof. Pap. 1515, US Geol. Surv., Washington DC, 1990).
27. McKay, A. in *Reports of Geological Explorations during 1888–1889* Vol. 20, 1–16 (NZ Geol. Surv., Wellington, 1890).
28. Cowan, H. The North Canterbury earthquake of September 1, 1888. *J. R. Soc. NZ* **21**, 1–12 (1991).
29. Eiby, G. An annotated list of New Zealand earthquakes, 1460–1965. *NZ J. Geol. Geophys.* **11**, 630–647 (1967).
30. Anderson, H., Webb, T. & Jackson, J. Focal mechanisms of large earthquakes in the South Island of New Zealand: implications for the accommodation of Pacific–Australia plate motion. *Geophys. J. Int.* **115**, 1032–1054 (1993).
31. Working Group on California Earthquake Probabilities *Probabilities of Large Earthquakes Occurring in California on the San Andreas Fault* (Open-File Rep. 88–398, US Geol. Surv., 1988).
32. Working Group on California Earthquake Probabilities. Seismic hazards in southern California: Probable earthquakes, 1994–2024. *Bull. Seismol. Soc. Am.* **85**, 379–439 (1995).
33. Ward, S. & Valensise, G. Progressive growth of San Clemente Island, California, by blind thrust faulting: implications for fault slip partitioning in the California continental borderland. *Geophys. J. Int.* **123**, 285–293 (1990).
34. Cowan, H. Late Quaternary displacements on the Hope fault at Glynn Wye, North Canterbury. *NZ J. Geol. Geophys.* **33**, 285–293 (1990).
35. Berryman, K. *et al.* The Alpine Fault, New Zealand: variation in Quaternary structural style and geomorphic expression. *Annales Tectonicae* **VI**, 126–163 (1992).
36. Knuefer, P. Estimating ages of late Quaternary stream terraces from analysis of weathering rinds and soils. *Geol. Soc. Am. Bull.* **100**, 1224–1236 (1984).
37. Knuefer, P. Temporal variations in the latest Quaternary slip across the Australian–Pacific boundary, northeastern South Island, New Zealand. *Tectonics* **11**, 449–464 (1992).
38. Van Dissen, R. & Yeats, R. Hope fault, Jordan thrust, and uplift of the seaward Kaikoura range, New Zealand. *Geology* **19**, 392–396 (1991).
39. Little, T., Grapes, R. & Berger, G. Late Quaternary strike-slip on the eastern part of the Awatere Fault, South Island, New Zealand. *Bull. Geol. Soc. Am.* (in the press).
40. England, P. & Wells, R. E. Neogene rotations and quasicontinuous deformation of the Pacific Northwest continental margin. *Geology* **19**, 978–981 (1991).
41. Nur, A. & Mavko, G. Postseismic viscoelastic rebound. *Science* **183**, 204–206 (1974).
42. Turcotte, D. & Spence, D. An analysis of strain accumulation on a strike slip fault. *J. Geophys. Res.* **79**, 4407–4412 (1974).
43. Prescott, W. & Yu, S.-B. Geodetic measurement of horizontal deformation in the northern San Francisco Bay region, California. *J. Geophys. Res.* **91**, 7475–7484 (1986).
44. Pearson, C., Beavan, J., Darby, D., Blick, G. & Walcott, R. Strain distribution across the Australian–Pacific plate boundary in the central South Island, New Zealand, from 1992 GPS and earlier terrestrial observations. *J. Geophys. Res.* **100**, 22071–22081 (1995).
45. Savage, J. Equivalent strike-slip earthquake cycles in half space and lithospheric-asthenosphere earth models. *J. Geophys. Res.* **95**, 4873–4879 (1990).
46. Savage, J. & Prescott, W. Asthenosphere readjustment and the earthquake cycle. *J. Geophys. Res.* **83**, 3369–3376 (1978).
47. Zoback, M. *et al.* New evidence on the state of stress on the San Andreas fault system. *Science* **238**, 1105–1111 (1987).
48. Gilbert, L., Scholz, C. & Beavan, J. Strain localisation along the San Andreas Fault; consequences for loading mechanisms. *J. Geophys. Res.* **12**, 23975–23984 (1994).
49. Ward, S. & Valensise, G. The Palos Verdes terraces, California: Bathtub rings from a buried reverse fault. *J. Geophys. Res.* **99**, 4485–4494 (1994).
50. McNeilan, T., Rockwell, T. & Resnik, G. Style and rate of Holocene slip, Palos Verdes Fault, Southern California. *J. Geophys. Res.* **101**, 8317–8334 (1996).
51. DeMets, C., Gordon, R., Argus, D. & Stein, S. Current plate motions. *Geophys. J. Int.* **101**, 425–478 (1990).
52. DeMets, C., Gordon, R., Argus, D. & Stein, S. The effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions. *Geophys. Res. Lett.* **21**, 2191–2194 (1994).
53. Molnar, P. & Gipson, J. Very long baseline interferometry and active rotations of crustal blocks in the Western Transverse Ranges, California. *Geol. Soc. Am. Bull.* **106**, 594–606 (1994).

Acknowledgements. We thank T. Arnadottir, G. Blick, G. Rowe, R. I. Walcott and P. Wood for their help in the field; J. Beavan, D. Darby and R. I. Walcott for discussions on the measurement and interpretation of active deformation in New Zealand; K. Sieh for many insights into the active deformation of California; Z.-k. Shen and D. D. Jackson for provision of data; and J. Haines for comments. GPS studies of crustal deformation in the South Island of New Zealand were supported at Oxford by the Natural Environment Research Council. The GPS measurements in the Marlborough fault zone were also supported by the Department of Surveying and Land Information, the Institute of Geological and Nuclear Sciences, and Victoria University, Wellington. Part of this work was carried out while P.C.E. held a visiting professorship at the California Institute of Technology.

Correspondence and requests for materials should be addressed to S.J.B. (e-mail: S.J.Bourne@siep.shell.com).

Crystal structure of the yeast MAT α 2/MCM1/DNA ternary complex

Song Tan & Timothy J. Richmond

ETH-Zürich, Institut für Molekularbiologie und Biophysik, ETH-Hönggerberg, CH-8093 Zürich, Switzerland

The structure of a complex containing the homeodomain repressor protein MAT α 2 and the MADS-box transcription factor MCM1 bound to DNA has been determined by X-ray crystallography at 2.25 Å resolution. It reveals the protein-protein interactions responsible for cooperative binding of MAT α 2 and MCM1 to DNA. The otherwise flexible amino-terminal extension of the MAT α 2 homeodomain forms a β -hairpin that grips the MCM1 surface through parallel β -strand hydrogen bonds and close-packed, predominantly hydrophobic, side chains. DNA bending induced by MCM1 brings the two proteins closer together, facilitating their interaction. An unusual feature of the complex is that an eight-amino-acid sequence adopts an α -helical conformation in one of two copies of the MAT α 2 monomer and a β -strand conformation in the other. This 'chameleon' sequence of MAT α 2 may be important for recognizing natural operator sites.

Specific recognition of eukaryotic gene regulatory sites often requires the collaboration of several DNA-binding transcription factors. For example, homeodomain transcription factors, which guide cell growth and development in most eukaryotes, have low DNA-binding specificity on their own and bind DNA in combination with other transcription factors to increase binding affinity and specificity¹⁻⁴. This approach also enables combinatorial control, or the modular combination of a limited number of factors to control expression of a variety of genes.

The homeodomain protein, MAT α 2 (abbreviated here as α 2), and the MADS-box protein MCM1 exemplify such combinatorial control in determining mating type in the yeast *Saccharomyces cerevisiae*⁵. In the haploid α -cell, α 2 interacts with the cell-type-independent transcription factor MCM1 to repress α -specific genes. In contrast, α 2 heterodimerizes with another homeodomain factor, MAT α 1, to repress haploid-specific genes in diploid cells. The 210-amino-acid α 2 protein contains two domains separated by a flexible linker⁶. The amino-terminal 100-amino-acid domain is required for repression and interacts with the TUP1 component of the general repression mediator complex TUP1/SSN6 (refs 7, 8). The 60-amino-acid homeodomain near the carboxy terminus binds DNA on its own without significant distortion of the DNA helix⁹. As with most other homeodomains characterized^{1,3}, the α 2 homeodomain positions the last of its three α -helices in the DNA major groove and inserts its N-terminal arm into the adjacent minor groove. The α 1/ α 2/DNA crystal structure shows that the otherwise unstructured α 2 C-terminal tail required for α 1 interaction⁹⁻¹¹ packs as an amphipathic helix against a hydrophobic surface on the α 1 homeodomain and that DNA bending by α 1/ α 2 is required for the interaction¹².

In addition to its role in mating type determination, the essential gene product MCM1 is required for diverse cellular processes such as cell-cycle control¹³, minichromosome maintenance¹⁴ and stress tolerance¹⁵. Besides associating with α 2, MCM1 binds with MAT α 1 to activate α -specific genes in α -cells, with STE12 to mediate pheromone response, and with SFF to control G2 phase-specific transcription of SWI5 and CLB2 (refs 16, 17). The 286-amino-acid MCM1 protein contains an 80-amino-acid domain near its amino terminus which is sufficient for viability, cell-type-specific transcription and minichromosome maintenance¹⁸. This same domain is highly conserved in the mammalian serum response factor (SRF) and specifies DNA binding, dimerization and interaction with accessory factors¹⁹. The 56-residue MADS-box, which is shared by

the mammalian myocyte enhancer factor 2 (MEF2) transcription factors and many plant homeotic genes such as *agamous* and *deficiens*²⁰, is located at the beginning of this 80-residue domain. Our crystal structure of the SRF core/DNA complex revealed that the MADS-box dimer binds to and bends DNA using a pair of antiparallel coiled-coil α -helices packed against an antiparallel β -sheet²¹.

MCM1 and α 2 bind cooperatively to the 30- or 31-base-pair (bp) α -specific operator DNA, which contains the pseudosymmetric MCM1 16-bp P-box-binding site flanked on either side by α 2-binding sites. An α 2 dimer binds on its own with equal affinity to such binding sites even when the distance between the sites is varied over an entire DNA helical turn or when the orientation of the sites is altered²². The linker between the two domains must be flexible to accommodate these differences, and both limited protease analysis and NMR studies suggest that the linker is disordered in the free protein^{6,23}. Binding in the presence of MCM1 increases the affinity of α 2 for DNA by 50- to 500-fold, but only when the spacing and orientation of the α 2 binding sites is limited to that found in naturally occurring operators^{22,24,25}. The cooperative binding is achieved by direct interaction between MCM1 and the flexible linker of α 2 (refs 23, 26).

We have determined the structural basis for the interaction between MCM1 and α 2 by X-ray crystallography. We have crystallized a ternary complex of the conserved MCM1 core with α 2 (103-189) bound to operator DNA and solved the structure of this complex to 2.25 Å resolution. The structure shows that α 2 uses a short eight-amino-acid module in the linker region to bind MCM1 via parallel β -strands. This recognition strand is tethered to the α 2 homeodomain by a segment of unusual conformational flexibility. Additional contacts are also made between the MAT α 2 homeodomain itself and MCM1. The complex further shows the structure of MCM1 and its interactions with DNA for the first time. Although MCM1 and SRF have the same overall fold and mode of DNA binding, significant differences in DNA recognition occur.

The α 2/MCM1/DNA complex

The complex shows a triangular slab of MCM1(1-100) dimer flanked by two copies of α 2 and bound to bent B-form DNA (Fig. 1a, b). The proteins cover the DNA fragment extensively, making hydrogen-bond or van der Waals contact with all but 2 of the 26 base pairs. Each MCM1 monomer contains the same three layers

observed in the SRF core: a 7.5-turn α -helix at the base, a β -hairpin in the middle and a coiled region followed by a 3-turn α -helix at the apex²¹. Dimerization occurs along the extensive flat side of the monomer. MCM1 interacts with DNA predominantly with its long α -helices located nearly parallel to the minor groove at the centre of the P-box. These α -helices extend into the major groove on either side of the dyad and direct contacts made within the major groove and along the phosphate backbone cause the DNA to bend around MCM1. The N-terminal arm of MCM1 passes over the DNA backbone in a manner similar to SRF, but the subsequent interactions with the DNA grooves are substantially different.

An $\alpha 2$ homeodomain core binds to the major groove 1.5 turns away from the dyad, with the DNA-binding helix axis oriented approximately parallel to the MCM1 two-fold axis. The N-terminal arm of the homeodomain inserts into the minor groove between the major grooves contacted by $\alpha 2$ and MCM1. The extended chain then turns 90° to rise up above the DNA before an additional 90° turn leads to an antiparallel β -hairpin. Parallel β -strands are made

with the outer strand of the adjacent MCM1 monomer by $\alpha 2$ residues 115–120 in this 'cis' interaction between the two proteins, extending the four-stranded MCM1 β -sheet to six strands. MCM1-induced bending of DNA is necessary to position $\alpha 2$ correctly for this interaction to occur.

Although the co-crystallization 26-bp DNA was designed to bind a dimer of MCM1 and only one monomer of $\alpha 2$ (103–189), a second copy of $\alpha 2$ (103–189) fortuitously binds at the junctions of the continuous DNA helix formed by crystal contacts (Fig. 1c). This homeodomain retains the same fold as its counterpart and binds DNA similarly, although the homeodomain N-terminal arm does not insert as deeply into the adjacent minor groove. The chain then turns by nearly 180° and adopts an α -helical conformation for 2.5 turns (residues 121–128) before converting abruptly into a β -strand (residues 115–120), similar in conformation to the first $\alpha 2$ copy. This β -strand makes a 'trans' interaction with a crystal-symmetry-related copy of MCM1 that is essentially identical to the 'cis' $\alpha 2$ /MCM1 interaction (Fig. 1a–c). Residues 121–128 of $\alpha 2$ comprise a

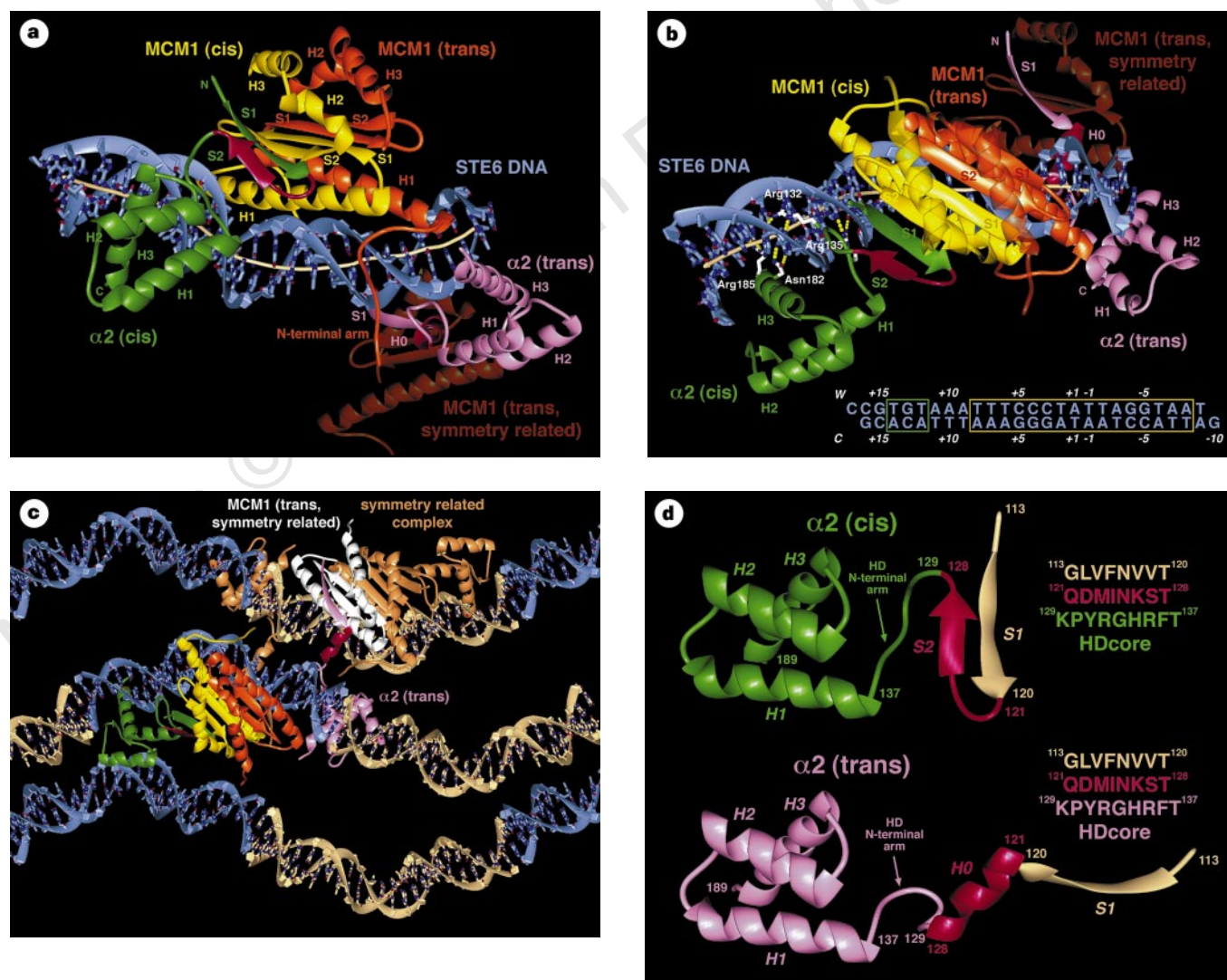


Figure 1 Structure of the $\alpha 2$ /MCM1/STE6 DNA complex. **a**, The ternary complex in ribbon representation viewed from the side. The DNA helical axis, calculated in CURVES⁴¹, is shown as a thin beige tube. $\alpha 2$ residues 121–128, which comprise the chameleon sequence, are shown in red. **b**, The ternary complex viewed from the top. The 16-bp P-box and the $\alpha 2$ core binding-site sequences are outlined in yellow and green respectively. **c**, Crystal packing interactions in the $\alpha 2$ /MCM1/DNA crystal. The base-pairing interaction at the ends of the DNA produces a continuous DNA helix in the crystal, with a head-to-tail packing of complexes

along this axis. The interactions between *trans* $\alpha 2$ (pink) and the symmetry-related *trans* MCM1 monomer (white) provides a second source of crystal interactions roughly perpendicular to the DNA helix axis. Despite its appearance, helix 2 of *cis* $\alpha 2$ (green) does not interact with the symmetry-related DNA (blue). **d**, Comparison of *cis* and *trans* $\alpha 2$ copies. The two $\alpha 2$ copies have been aligned by the α -carbon atoms of residues 138–187 within the homeodomain core (r.m.s.d., 0.335 Å). The homeodomain is shown in green or pink, the chameleon sequence in red, and strand 1, which interacts with MCM1, in beige.

'chameleon' sequence, since this region is capable of adopting different secondary structures (α -helix or β -strand; Fig. 1d) in different environments²⁷.

MCM1-DNA and α 2-DNA interactions

MCM1 and SRF employ the same structural elements of the N-terminal arm, the long α -helix and the β -hairpin loop to interact with DNA. The α -helix is responsible for most of the DNA contacts and interacts with both the major groove and phosphate backbone through hydrogen bonds and hydrophobic contacts (Fig. 2a). The interactions of the MCM1 and SRF α -helices with DNA within the central 10-bp core DNA site are generally conserved. In particular, the insertion of Lys 38 in the major groove to hydrogen bond to the O6 and N7 groups of adjacent G residues at positions ± 5 and ± 4 , respectively, is maintained in the two complexes. These interactions account for the requirement of guanine at position ± 5 and the relaxed requirement of a purine at ± 4 observed in mutagenesis and

DNA selection experiments using MCM1 (refs 28, 29). Other DNA interactions made by both SRF and MCM1 using conserved MADS-box family residues include the triad of Arg 39, Lys 45 and Lys 46, which hydrogen-bond to the phosphate backbone of the very narrow minor groove near the dyad (Fig. 2a), and the close contact made by Gly 42 with the phosphate backbone in the same region.

MCM1, like SRF, uses its long α -helices to bend DNA. The SRF/DNA and α 2/MCM1/DNA complexes exhibit similar DNA bends of about 72°, although distributed differently: MCM1 bends the DNA by 37° on the *cis* α 2 proximal edge of the core 10-bp binding site, 19° within the central 6 bp, and only 16° on the *trans* α 2 proximal edge compared with 32, 15 and 27°, respectively, in the SRF/DNA complex. In contrast to the SRF/DNA complex, MCM1 also introduces an additional bend, roughly perpendicular to the first, around position +8 (Figs 1b, 2b). The second bend of about 20° brings the adjacent *cis* α 2 binding site closer to MCM1 to allow

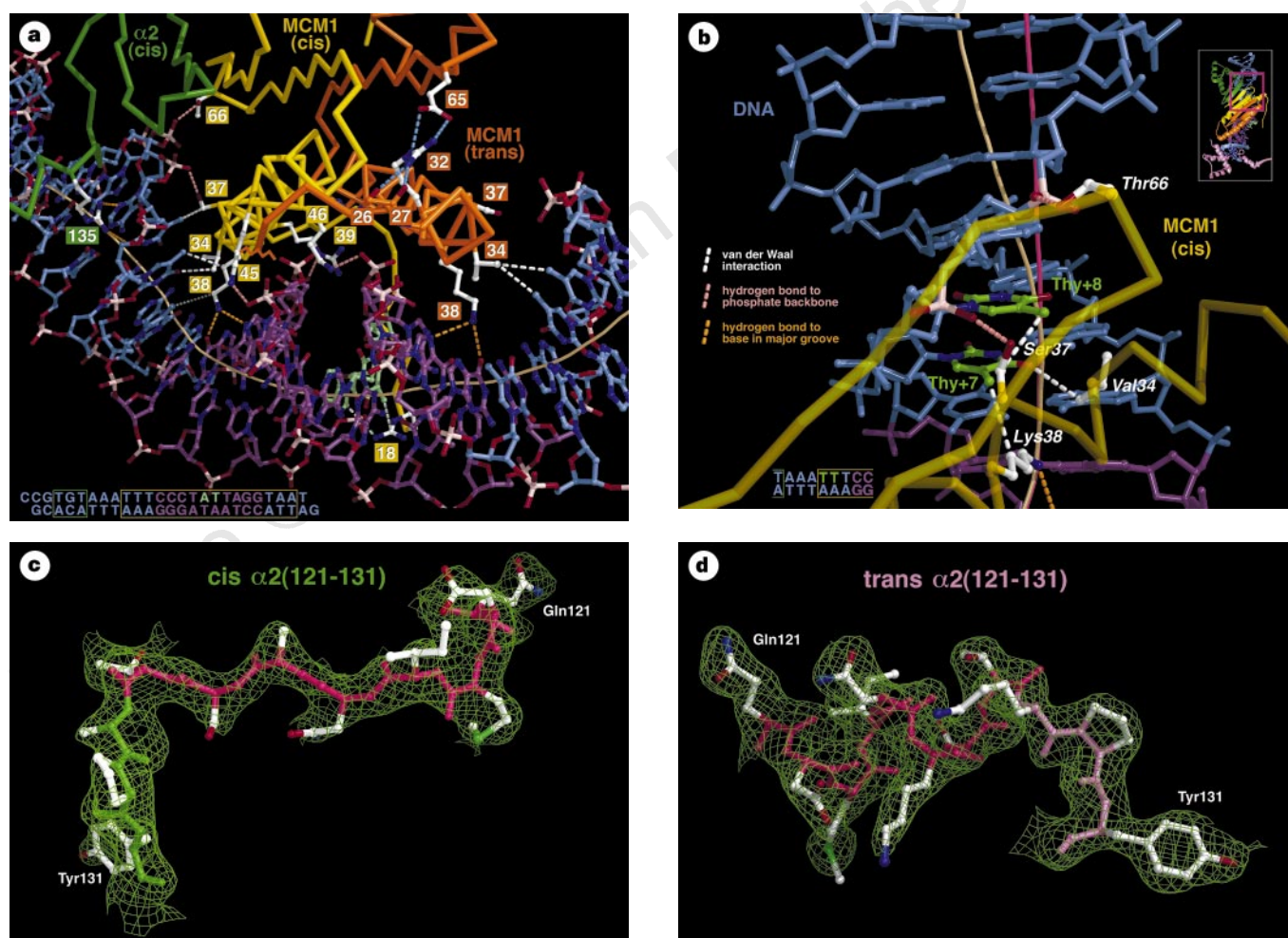


Figure 2 Details of MCM1-DNA interactions and electron density of chameleon sequences. **a**, MCM1-DNA interactions viewed approximately 15° from MCM1 dimer interface plane. Rods joining α -carbon positions of the protein backbone are shown, except for *trans* MCM1 monomer residues 26 and 27 (all backbone atoms shown) and residues 18–23 (omitted not to obscure other interactions). Hydrogen bonds between protein and DNA bases (orange), protein and DNA backbone (pink) and within MCM1 (light blue) are shown. Van der Waals interactions are represented in white, but an additional possible contact between *cis* MCM1 Lys 38 and Thy+6 (4.26 Å) is displayed with grey dots. **b**, Source of the DNA bend away from *cis* MCM1 helix 1. The helix axis is shown in beige; the helix axis for DNA extended from base pair +7 as straight B-form is shown in dark pink. **c**, Electron density ($2F_o - F_c$) contoured at 1.0 σ for the *cis* α 2 region including the

β -chameleon structure. The electron density for the β -chameleon structure is the poorest in the complex, with the possible exception of the terminal residues, consistent with the higher B-factors and presumed relative flexibility and mobility of this region. No electron density is observed for the side chains of Lys 126 and Lys 129 and these have been modelled as alanine residues. A simulated annealing omit map for *cis* α 2(121–128), contoured at 0.6 σ , yields a slightly noisier version with continuous electron density evident for the main and side chains shown. **d**, Electron density ($2F_o - F_c$) contoured at 1.3 σ for the *trans* α 2 region including the α -chameleon structure. A simulated annealing omit map for *trans* α 2(121–128) contoured at 0.8 σ yields a slightly noisier version, but continuous electron density is evident for the main and side chains shown.

$\alpha 2$ /MCM1 interaction. The DNA bends are induced by MCM1–DNA hydrophobic interactions in the major groove, by the base-specific contacts of Lys 38, and by hydrogen bonds to the phosphate backbone.

Hydrophobic interactions by MCM1 with 5-methyl groups of the thymine-rich sequence flanking the core MCM1 site are critical in determining the DNA path, especially on the *cis* side. The first DNA bend, contributed in part by van der Waals contacts by Thy+7 and possibly Thy+6 with Val 34 and Lys 38, brings Ser 37 unacceptably close to the phosphate backbone. The DNA bends away from its otherwise straight path to avoid this close contact, resulting in the second bend. The altered path of the DNA is stabilized with van der Waals contact between Ser 37 β -carbon and Thy+8 5-methyl group, and with phosphate contacts by Ser 37 and Thr 66 in the sole DNA contact made by the MCM1 β -sheet layer (Fig. 2b). On the *trans* side, the shallower bend is accompanied by only Val 34 making van der Waals contacts with the 5-methyl groups of thymine at positions –7 and –8 (Fig. 2a). Ser 37 is now positioned too far from the DNA to make any contact and the second bend is absent. By comparison, the SRF/DNA complex makes less intimate contact at positions +6 and +7 because there are no thymine residues with hydrophobic 5-methyl groups at these positions. The resulting DNA path allows van der Waals contact of Ser 37 and Thy+8 without requiring a second bend.

These structural results compare favourably with mutagenesis experiments which have shown that thymines at positions +7 and +8 cannot be substituted without substantially reducing both MCM1-mediated activation and $\alpha 2$ /MCM1-mediated repression^{29,30}. Substitution of guanine for thymine at +7 also reduces DNA bending by MCM1, even though binding affinity is not significantly affected²⁹. This is consistent with the hydrophobic contacts by Val 34 and Lys 38 with Thy+7 that are used to distort the DNA helical axis, and it underlines the importance of DNA bending for MCM1 activity.

As observed in the SRF/DNA complex²¹, Arg 32 is essential in establishing the path of the MCM1 N-terminal extension across to the adjacent minor groove by forming hydrogen bonds with Glu 65 in the β -loop and with the main-chain oxygen atoms of Ile 26 and Glu 27 simultaneously (Fig. 2a). However, the paths of the N-terminal extensions from the two MCM1 monomers and from SRF diverge when they approach the minor groove. The *cis* MCM1 monomer hydrogen-bonds with the phosphate oxygens of Ade+1 and Thy–1 and uses the side chain of Arg 18 to make van der Waals contacts with the C5 methyl group of Thy–1 and the C8 atom of Ade+1 in the major groove (Fig. 2a). Contacts in the major groove at the centre of the MCM1 binding site have been predicted based on crosslinking and carboxyethylation results^{28,29}. No minor groove contacts are made by *cis* MCM1, in contrast to the extensive interaction of SRF's Arg 143 (corresponds to MCM1 Arg 18) in the minor groove²¹. The *trans* copy of MCM1 does make a minor-groove hydrogen bond to O2 of Thy–3 through Arg 19 in what

appears to be a consequence of crystal packing because the interaction is stabilized by *trans* $\alpha 2$ Tyr 131, in contrast to the *cis* MCM1 N-terminal arm which is not constrained by crystal packing.

No interpretable electron density for MCM1 *cis* residues 2–14 and *trans* residues 2–17 was observed, and we presume that these residues are disordered in the crystal. These N-terminal residues are not necessary for cell-type-specific transcription or viability, but appear to be involved in a phosphorylation-dependent manner in the yeast cell's response to stress, for example high salt concentrations¹⁵.

The structure of the *cis* $\alpha 2$ homeodomain, the conformation of its DNA-binding site, and the interactions between $\alpha 2$ and DNA are, with few exceptions, unchanged from the $\alpha 2$ /DNA⁹ and $\alpha 1/\alpha 2$ /DNA complexes¹². All five copies of $\alpha 2$ present in the two ternary and one binary complexes share the same homeodomain globular portion, with only minor variations in the loops connecting the α -helices. The conformation of DNA contacted by the $\alpha 2$ homeodomains is also similar in the three complexes. However, the network of water molecules in the $\alpha 1/\alpha 2$ /DNA complex between $\alpha 2$ and DNA, including those waters mediating hydrogen bonds by Ser 181, is not observed in the $\alpha 2$ /MCM1/DNA complex. It is possible that either the lack of a phosphate group on the terminal nucleotide of the synthetic oligonucleotide has prevented the water network from forming³¹ or that higher-resolution data will be necessary to observe water molecules in this region.

$\alpha 2$ -MCM1 interactions (*cis* copy)

The *cis* copy of the $\alpha 2$ flexible linker implicated in the interaction with MCM1 (refs 23, 26) forms a β -hairpin with a type-I turn. The N-terminal strand of this $\alpha 2$ β -hairpin pairs through parallel strands with β -strand 2 of the adjacent MCM1 monomer, complemented by close packing of $\alpha 2$ and MCM1 side chains (Fig. 3a, c). With the exception of Val 119, each of the $\alpha 2$ residues between Leu 114 and Gln 121 (LVFNVVTQ) makes van der Waals contacts with MCM1 residues. The close packing at the $\alpha 2$ /MCM1 interface correlates well with mutagenesis experiments in which the identity of each $\alpha 2$ residue between 114 and 121 was shown to be important for $\alpha 2$ -dependent repression²⁶. Furthermore, disrupting the close packing by the MCM1 mutation S73R reduces the ability of MCM1 to bind $\alpha 2$ in the ternary complex without affecting binary MCM1/DNA complex formation³². Each interaction of *cis* and *trans* $\alpha 2$ (114–121) with MCM1 buries 1,200 Å² of accessible surface area, compared to 3,200 Å² and 6,400 Å² for the MCM1/DNA and *cis* $\alpha 2$ /MCM1/DNA complexes, respectively.

At either end of the parallel strands, $\alpha 2$ residues interact with a hydrophobic pocket or groove created by the junction of several MCM1 secondary structure elements (Fig. 3a). These interactions are facilitated by the right-handed twist of the β -sheet formed by the $\alpha 2$ and MCM1 β -hairpins. At the N-terminal end, the large hydrophobic phenyl group of Phe 116 inserts into a hydrophobic

Table 1 Crystallographic analysis

	Native 1	Native 2	
Diffraction data			
X-ray source	Synchrotron	Rotating anode	
Resolution (Å)	25–2.25	25–2.95	
Reflections (observations/unique)	327,727/35,613	148,806/16,698	
I/ σ -I	9.8	6.2	
I/ σ -I (final shell)	1.7 (2.29–2.25 Å)	2.4 (3.07–2.95 Å)	
Completeness (%)	94.6	99.0	
R_{merge}^*	6.5	10.2	
Refinement statistics			
Resolution (Å)	25–2.25	Mean B-factor (Å ²)	52.5
R-factor/free R-factor (%)	24.1/28.5	R.m.s. dev. bond length (Å)	0.006
Reflections, I > 2 σ (working/test)	34,859/1,846	R.m.s. dev. bond angle	1.16
Number of protein or DNA atoms	3,604	Number of water molecules	53

* $\sum_{hkl} \sum_i |I_{hkl,i} - \bar{I}_{hkl}| / \sum_{hkl} \sum_i I_{hkl,i}$, where $\bar{I}_{hkl}$ is the mean of measurements for a single hkl .

pocket created by MCM1 residues Val 69, Val 81, Arg 87 and Ile 90 from the strand 2, coil (helix 2) and helix 3 segments, and capped on the side by $\alpha 2$ residue Leu 114 (Fig. 3a, c). MCM1 Arg 87, which forms the roof of the hydrophobic pocket, is held in position by a hydrogen bond to MCM1 Val 81 main chain. Phe 116 has been identified by alanine scanning mutagenesis as being essential for $\alpha 2$ binding to MCM1, and the F116A mutant was the most severely impaired of the mutants in the region, with only ~5% of wild-type $\alpha 2$ -dependent repression activity remaining²⁶. The MCM1 mutant V81F is defective in cooperative binding with $\alpha 2$ and DNA, presumably because the bulkier side chain partially fills the hydrophobic pocket otherwise occupied by $\alpha 2$ Phe 116³².

At the C-terminal end of the parallel strands, the aliphatic portion of $\alpha 2$ Gln 121 side chain lies along a hydrophobic groove of MCM1 to allow hydrogen-bond formation between the glutamine amide group and MCM1 Ser 51 (Fig. 3a, c). The hydrophobic groove is formed by residues Phe 48 and Val 52 from helix 1, Phe 72 from strand 2, and Thr 74 and Pro 75 from the coiled region of MCM1. SRF and MCM1 follow the same chain direction in the junction between strand 2 and the coiled region, despite the Arg-to-Pro change at residue 75, and MCM1 uses the imino group of Pro 75 to make van der Waals contact with $\alpha 2$ Gln 121.

In addition to these interactions at the parallel strand interface, hydrogen-bond and van der Waals contacts are observed between the $\alpha 2$ homeodomain itself and MCM1. Both contacts involve

residues in the N-terminal arm within the $\alpha 2$ homeodomain and β -strand 2 of MCM1. MCM1 Tyr 70 reaches below $\alpha 2$ strand 2 and makes van der Waals contact with $\alpha 2$ Asn 117 before forming a hydrogen bond with $\alpha 2$ His 134 in the N-terminal homeodomain arm (Fig. 3a). This conformation of the His 134 side chain requires a repositioning of $\alpha 2$ Tyr 131 residue as compared to the $\alpha 1/\alpha 2$ /DNA complex (Tyr 131 was not observed in the $\alpha 2$ /DNA complex). The repositioning eliminates a hydrogen bond between Tyr 131 and DNA present in the $\alpha 1/\alpha 2$ /DNA complex, but allows instead van der Waals contact between Tyr 131 and MCM1 Leu 68.

$\alpha 2$ -MCM1 interactions (*trans* copy)

In contrast to the β -hairpin present in the *cis* $\alpha 2$ copy, the linker of the *trans* copy forms a β -strand followed by an α -helix. The β -strand is very similar in conformation to the N-terminal strand of the *cis* $\alpha 2$ β -hairpin (0.38 Å r.m.s.d. for backbone atoms) and makes essentially the same interactions with MCM1 as *cis* $\alpha 2$, except that *trans* $\alpha 2$ binds a crystal-symmetry-related MCM1 monomer (Fig. 3a–d). The α -helix formed by *trans* $\alpha 2$ residues Gln 121 to Thr 128 projects away from the *trans* $\alpha 2$ homeodomain to bridge the distance to the symmetry-related MCM1 (Fig. 1c), preventing any interaction between $\alpha 2$ homeodomain residues His 134 and Tyr 131 with MCM1. The finding that *trans* $\alpha 2$ binds MCM1 with essentially identical interactions as *cis* $\alpha 2$, even though this requires a 90° change in chain direction and a displacement of 21 Å (between

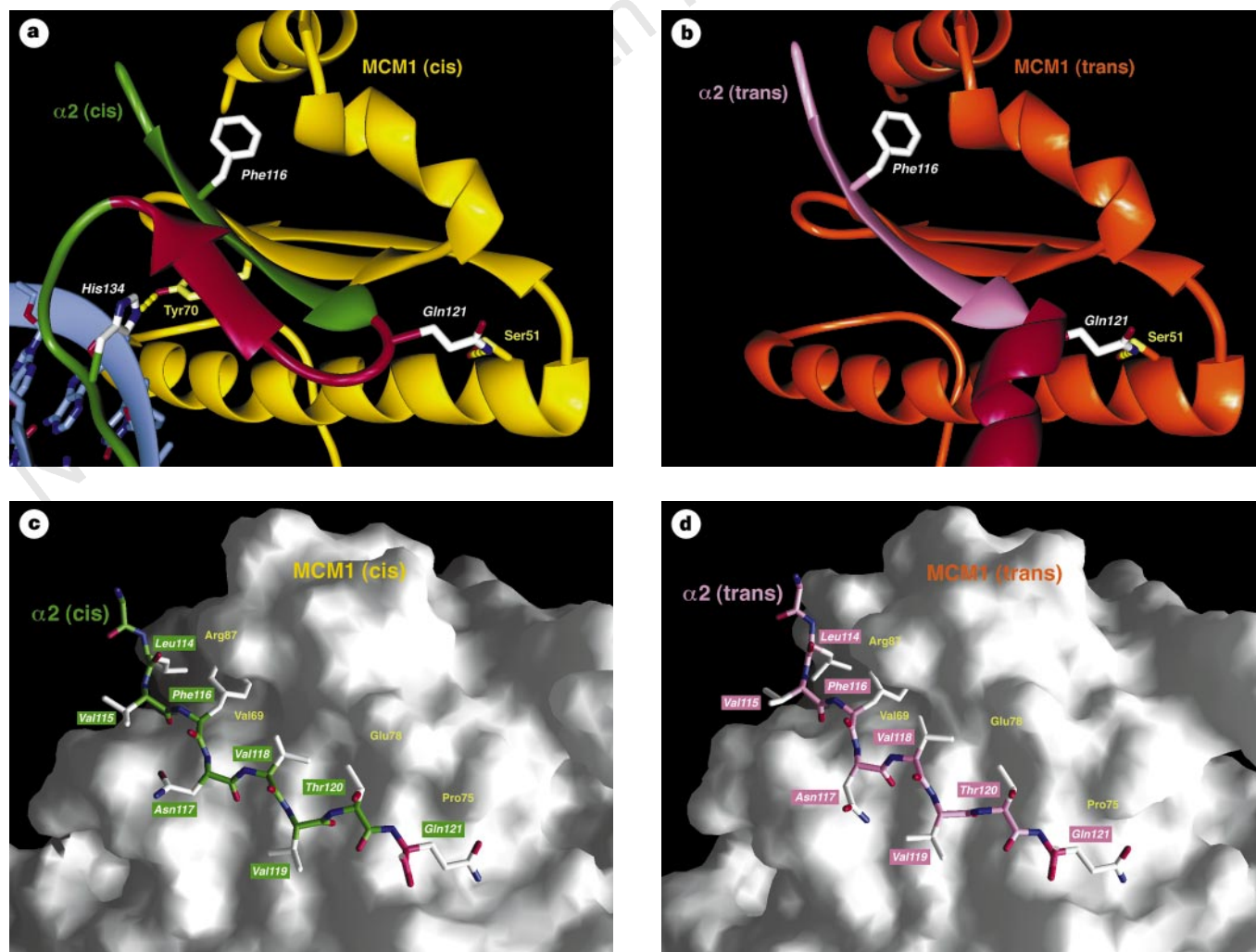


Figure 3 $\alpha 2$ -MCM1 interactions. **a, b**, *Cis* and *trans* $\alpha 2$ /MCM1 interactions in ribbon representation; selected side chains are shown. **c, d**, Surface representation of *cis* and *trans* $\alpha 2$ /MCM1 interactions showing the complementarity of $\alpha 2$

and MCM1 surfaces (prepared in GRASP⁴⁰). The view is slightly rotated compared with that in **a** and **b** in order to show better the hydrophobic cavity filled by $\alpha 2$ Phe 116 and hydrophobic channel occupied by Gln 121.

α -carbon atoms of *cis* and *trans* residue 114), suggests that $\alpha 2$ (114–121) has significant affinity for its MCM1 binding site and that $\alpha 2$ (121–134) has considerable conformational flexibility.

A chameleon sequence in the $\alpha 2$ linker

The $\alpha 2$ residues 121–128, QDMINKST, have the unusual property of forming either an α -helix or a β -strand. The β -strand formed in the *cis* $\alpha 2$ molecule makes four main-chain hydrogen bonds to its antiparallel counterpart to form the β -hairpin. The residues in this β -chameleon structure are significantly more mobile than other parts of the entire complex, with backbone atom B-factors averaging 83, as compared to 52 for the whole structure. The side chains are also very mobile: no discernible electron side-chain density is observed for Lys 126, and only weak side-chain density is present for Met 123 and Ile 124 (Fig. 2c). The mobility of the chameleon β -strand probably reflects its lack of interaction with other parts of the complex. No direct interactions are observed either for the hydrogen-bonding main-chain atoms on the solvent-accessible side of the chameleon β -strand or for the side-chain atoms, with the exception of Asp 122, which makes a hydrogen bond to the main-chain nitrogen of Ile 124. In contrast, the helix adopted by the same sequence in the *trans* $\alpha 2$ molecule is less mobile (average *B* of 51) and well defined electron density is present for most of the side chains (Fig. 2d). Each of the eight residues in the chameleon α -helix occupies the core α -helical region in a Ramachandran plot.

The sequence QDMINKST is a naturally occurring chameleon sequence characterized structurally in its two states under identical conditions since they occur in the same crystal. The term 'chameleon' was coined to describe a bistable sequence by Minor and Kim who engineered an 11-residue peptide sequence to fold as either an α -helix or β -strand in appropriate locations of the protein G IgG-binding domain²⁷. Other chameleon structures occur in the GDP- and GTP-bound protein synthesis elongation factor EF-Tu^{33,34} and the serpin family of serine proteinase inhibitors before and after cleavage³⁵. Chameleon sequences may be important in certain human and animal diseases. A bistable peptide sequence that normally adopts an α -helical form might aggregate as β -strands to form amyloid fibres if the sequence is induced to switch to the β -conformation: such a mechanism may implicate chameleon

sequences in amyloid fibre diseases such as Alzheimer's disease and bovine spongiform encephalopathy (BSE)³⁶.

Model for $\alpha 2$ -MCM1 interaction on 31-bp site

Of the five natural *a*-specific promoters, four contain different spacings between the flanking $\alpha 2$ and central MCM1-binding sites. The central 16-bp MCM1-binding sites within the STE6, STE2, MFA1 and MFA2 operators are separated from the $\alpha 2$ -binding sites by 2 bp on one side and 3 bp on the other. Thus, in addition to the 3-bp spacing observed in the crystal, $\alpha 2$ must also bind to a DNA site 1 bp closer to MCM1 while still interacting with MCM1 through the flexible linker. Given the *trans* $\alpha 2$ -MCM1 interaction we have observed in the crystal, it is likely that $\alpha 2$ bound to a 2-bp spaced site will interact with MCM1 using residues Leu 114 to Thr 120, and possibly Gln 121, in the same way as in the *cis* (3-bp spacing) interaction. The $\alpha 2$ homeodomain-DNA interactions are also likely to remain unchanged between the 2- or 3-bp-spaced sites because the sequences of all $\alpha 2$ binding sites are very similar and not affected by the spacing between the binding sites. Therefore, $\alpha 2$ binding to the 2- and 3-bp-spaced sites will probably differ mainly in the conformation of residues 122–134 between the $\alpha 2$ homeodomain and its recognition β -strand. The *cis* $\alpha 2$ conformation is not compatible with the 2-bp spacing because severe distortion of the DNA would be required to position the $\alpha 2$ site appropriately.

The conformation of $\alpha 2$ (122–134) bound to a 2-bp-spaced site will depend on the MCM1-induced bend in the DNA as this determines the precise location of the $\alpha 2$ binding site and therefore of $\alpha 2$ itself. If MCM1 bends the DNA in solution as observed on the *trans* side of the complex, the shallower bend places $\alpha 2$ significantly further away from MCM1 even with a 2-bp spacing. Interaction between $\alpha 2$ strand 1 and MCM1 would be possible only if $\alpha 2$ (122–134) were in an extended configuration to cover the ~ 30 Å distance. However, we suspect that MCM1 bends *a*-specific upstream activating sequence (UAS) DNA symmetrically in solution and that the shallower bend on the *trans* side is induced in the crystal by the binding of *trans* $\alpha 2$ (*trans* $\alpha 2$, which is responsible for important crystal contacts, would be occluded if MCM1 bent the DNA sharply on the *trans* side). Model building suggests that the conformation of the $\alpha 2$ homeodomain N-terminal arm could be maintained up to and including Pro 130, preserving the minor groove contacts made by Arg 132, Gly 133 and Arg 135; $\alpha 2$ (122–129) could then bridge to the recognition β -strand using a short α -helical segment similar to that in the *trans* $\alpha 2$ copy (Fig. 4). The hydrogen bond between $\alpha 2$ His 134 and MCM1 Tyr 70 in the 3-bp-spaced site interaction could be substituted in this 2-bp-spaced site model by His 134 to DNA-phosphate contacts; the hydrophobic contact of $\alpha 2$ Tyr 131 with MCM1 Leu 68 could be replaced by hydrogen bonds between $\alpha 2$ Tyr 131 and $\alpha 2$ Asn 117 or MCM1 Tyr 70.

Discussion

The structure of an $\alpha 2$ /MCM1/DNA complex shows that a short eight-residue module of $\alpha 2$ mediates the cooperative interaction between $\alpha 2$ and MCM1. In the absence of MCM1, this region is part of an unstructured, flexible linker that enables the $\alpha 2$ dimer to bind on its own *in vitro* to variously spaced sites with similar affinity. The spacing between $\alpha 2$ sites in naturally occurring promoters shows only a one-base-pair variability, reflecting the spatial interaction with MCM1 and possibly the involvement of additional secondary structure between the $\alpha 2$ homeodomain and the strand formed by the eight-residue module. We observe a second strand that pairs with the module strand in the 3-bp-spaced interaction in the crystal. Even though this second $\alpha 2$ strand does not interact directly with MCM1, it may be important indirectly because its formation will affect the energetics of binding between $\alpha 2$ and MCM1 on DNA. Thermodynamic studies should help to define the role of these direct and indirect induced-fit interactions in $\alpha 2$ /MCM1/DNA ternary complex formation. □

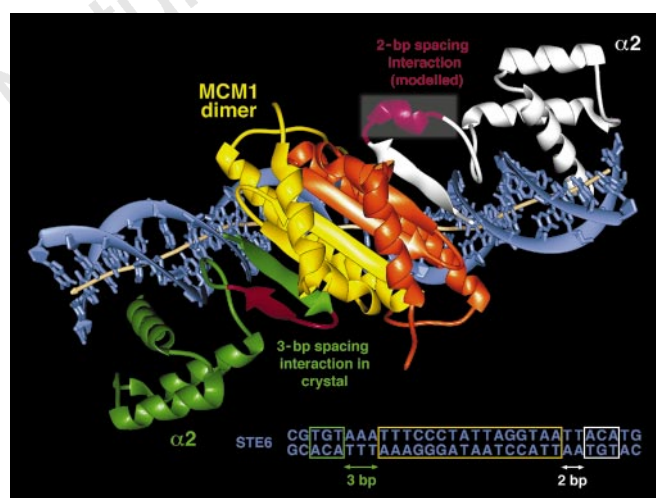


Figure 4 Model for the interaction of $\alpha 2$ and MCM1 on 31 bp of STE6 UAS. The β -chameleon structure and the equivalent region in the 2-bp-spaced $\alpha 2$ model are shown in red. The $\alpha 2$ /MCM1/STE6 DNA model is based on the MCM1 dimer, *cis* (3-bp-spaced) $\alpha 2$ and respective DNA coordinates from the crystal structure. The 2-bp-spaced $\alpha 2$ and remaining DNA were modelled by rotating the equivalent fragments from the crystal structure about the P-box pseudodyad axis before translating and rotating to account for the 1-bp shift of the $\alpha 2$ -binding site.

Methods

Data collection and processing. Purification and crystallization of the complex containing MCM1(1–100), $\alpha 2$ (103–189) and synthetic oligonucleotide fragments will be described elsewhere (manuscript in preparation). The crystals grew in space group $P2_12_12_1$, with cell dimensions $70.6 \times 72.6 \times 150.7$ Å and one complex containing one copy of DNA and two copies each of MCM1 and $\alpha 2$ per asymmetric unit. Crystals were prepared for cooling by transferring stepwise into stabilization buffer (50 mM sodium acetate, pH 5.5, 100 mM ammonium acetate, 10 mM CaCl_2 , 11.0% PEG 6000) containing 5, 10, 15, 20 and 25% glycerol. Individual crystals were mounted in polyethylene fibre loops and flashed-cooled by rapid immersion in liquid nitrogen. Diffraction data summarized in Table 1 were collected on an MAR image plate using either a Rigaku FR-C rotating-anode X-ray generator or synchrotron radiation on beamline BM14 at ESRF Grenoble, and processed using local programs (T.J.R., unpublished).

Structure determination and refinement. Molecular replacement was performed in AMoRe³⁷ using a model containing one copy of $\alpha 2$, a dimer of MCM1, and 24 base pairs of DNA based on SRF/DNA (PDB accession number, 1SR5) and $\alpha 2$ /DNA coordinates (PDB accession number, 1APL). The second highest rotation/translation solution was consistent with two iodine sites determined from a heavy atom derivative containing iodinated nucleotides. Rigid body refinement in XPLOR³⁸ allowing independent translation and rotation of MCM1/DNA and $\alpha 2$ /DNA components, followed by a single round of simulated annealing, improved the R_{free} factor from 56.9 to 54.6 to 50.1%. The resulting electron density showed clear density for the missing DNA nucleotides. Iterative rounds of model building using program O and refinement using XPLOR 3.843 (incorporating a bulk solvent model and anisotropic scaling for diffraction data) or TNT³⁹ were used to improve the model. Inclusion of lower-resolution data missing in the 2.25 Å data set improved the connectivity of the electron density map and established the presence of the second (*trans*) copy of $\alpha 2$. The final model contains *cis* MCM1(15–99), *trans* MCM1(18–98), *cis* $\alpha 2$ (113–189), *trans* $\alpha 2$ (113–189) and the 26-bp DNA. No interpretable density was observed for $\alpha 2$ residues 103–112 in either copy. A Ramachandran plot shows 92.1% of residues in the core allowed regions and no residues in the disallowed regions.

Figures 1, 2, 3a, b and 4 were produced in MIDAS⁴⁰.

Received 20 August; accepted 12 November 1997.

- Gehring, W. J. *et al.* Homeodomain-DNA recognition. *Cell* **78**, 211–223 (1994).
- Wilson, D. S. & Desplan, C. Homeodomain proteins. Cooperating to be different. *Curr. Biol.* **5**, 32–34 (1995).
- Wolberger, C. Homeodomain interactions. *Curr. Opin. Struct. Biol.* **6**, 62–68 (1996).
- Vershon, A. K. Protein interactions of homeodomain proteins. *Curr. Opin. Biotechnol.* **7**, 392–396 (1996).
- Johnson, A. D. Molecular mechanisms of cell-type determination in budding yeast. *Curr. Opin. Genet. Dev.* **5**, 552–558 (1995).
- Sauer, R. T., Smith, D. L. & Johnson, A. D. Flexibility of the yeast alpha 2 repressor enables it to occupy the ends of its operator, leaving the center free. *Genes Dev.* **2**, 807–816 (1988).
- Hall, M. N. & Johnson, A. D. Homeodomain of the yeast repressor alpha 2 is a sequence-specific DNA-binding domain but is not sufficient for repression. *Science* **237**, 1007–1012 (1987).
- Komachi, K., Redd, M. J. & Johnson, A. D. The WD repeats of Tup1 interact with the homeodomain protein alpha 2. *Genes Dev.* **8**, 2857–2867 (1994).
- Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. Crystal structure of a MATalpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* **67**, 517–528 (1991).
- Phillips, C. L., Vershon, A. K., Johnson, A. D. & Dahlquist, F. W. Secondary structure of the homeodomain of yeast alpha 2 repressor determined by NMR spectroscopy. *Genes Dev.* **5**, 764–772 (1991).
- Mak, A. & Johnson, A. D. The carboxy-terminal tail of the homeodomain protein alpha 2 is required for function with a second homeodomain protein. *Genes Dev.* **7**, 1862–1870 (1993).

- Li, T., Stark, M. R., Johnson, A. D. & Wolberger, C. Crystal structure of the MATalpha 1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science* **270**, 262–269 (1995).
- Lydall, D., Ammerer, G. & Nasmyth, K. A new role for MCM1 in yeast: cell cycle regulation of SW15 transcription. *Genes Dev.* **5**, 2405–2519 (1991).
- Passmore, S., Maine, G. T., Elble, R., Christ, C. & Tye, B. K. *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. *J. Mol. Biol.* **204**, 593–606 (1988).
- Kuo, M. H., Nadeau, E. T. & Grayhack, E. J. Multiple phosphorylated forms of the *Saccharomyces cerevisiae* Mcm1 protein include an isoform induced in response to high salt concentrations. *Mol. Cell. Biol.* **17**, 819–832 (1997).
- Althoefer, H., Schleiffer, A., Wassmann, K., Nordheim, A. & Ammerer, G. Mcm1 is required to coordinate G2-specific transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 5917–5928 (1995).
- Maher, M., Cong, F., Kindelberger, D., Nasmyth, K. & Dalton, S. Cell cycle-regulated transcription of the CLB2 gene is dependent on Mcm1 and a ternary complex factor. *Mol. Cell. Biol.* **15**, 3129–3137 (1995).
- Christ, C. & Tye, B. K. Functional domains of the yeast transcription/replication factor MCM1. *Genes Dev.* **5**, 751–763 (1991).
- Primig, M., Winkler, H. & Ammerer, G. The DNA binding and oligomerization domain of MCM1 is sufficient for its interaction with other regulatory proteins. *EMBO J.* **10**, 4209–4218 (1991).
- Shore, P. & Sharrocks, A. D. The MADS-box family of transcription factors. *Eur. J. Biochem.* **229**, 1–13 (1995).
- Pellegrini, L., Tan, S. & Richmond, T. J. Structure of serum response factor core bound to DNA. *Nature* **376**, 490–498 (1995).
- Smith, D. L. & Johnson, A. D. A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an alpha 2 dimer. *Cell* **68**, 133–142 (1992).
- Vershon, A. K. & Johnson, A. D. A short, disordered protein region mediates interactions between the homeodomain of the yeast alpha 2 protein and the MCM1 protein. *Cell* **72**, 105–112 (1993).
- Keleher, C. A., Goutte, C. & Johnson, A. D. The yeast cell-type specific repressor alpha 2 acts cooperatively with a non-cell-type-specific protein. *Cell* **53**, 927–936 (1988).
- Keleher, C. A., Passmore, S. & Johnson, A. D. Yeast repressor alpha 2 binds to its operator cooperatively with yeast protein Mcm1. *Mol. Cell. Biol.* **9**, 5228–5230 (1989).
- Mead, J., Zhong, H., Acton, T. B. & Vershon, A. K. The yeast alpha2 and Mcm1 proteins interact through a region similar to a motif found in homeodomain proteins of higher eukaryotes. *Mol. Cell. Biol.* **16**, 2135–2143 (1996).
- Minor, D. L. Jr & Kim, P. S. Context-dependent secondary structure formation of a designed protein sequence. *Nature* **380**, 730–734 (1996).
- Wynne, J. & Treisman, R. SRF and MCM1 have related but distinct DNA binding specificities. *Nucleic Acids Res.* **20**, 3297–3303 (1992).
- Acton, T. B., Zhong, H. & Vershon, A. K. DNA-binding specificity of Mcm1: operator mutations that alter DNA-bending and transcriptional activities by a MADS box protein. *Mol. Cell. Biol.* **17**, 1881–1889 (1997).
- Zhong, H. L. & Vershon, A. K. The yeast homeodomain protein Mat alpha 2 shows extended DNA binding specificity in complex with Mcm1. *J. Biol. Chem.* **272**, 8402–8409 (1997).
- Labeets, L. A. & Weiss, M. A. Electrostatics and hydration at the homeodomain DNA interface: chemical probes of an interfacial water cavity. *J. Mol. Biol.* **269**, 113–128 (1997).
- Bruhn, L. & Sprague, G. F. Jr MCM1 point mutants deficient in expression of alpha-specific genes: residues important for interaction with alpha 1. *Mol. Cell. Biol.* **14**, 2534–2544 (1994).
- Abel, K., Yoder, M. D., Hilgenfeld, R. & Jurnak, F. An alpha to beta conformational switch in EF-Tu. *Structure* **4**, 1153–1159 (1996).
- Polekhina, G. *et al.* Helix unwinding in the effector region of elongation factor EF-Tu-GDP. *Structure* **4**, 1141–1151 (1996).
- Wright, H. T. The structural puzzle of how serpin serine proteinase inhibitors work. *BioEssays* **18**, 453–464 (1996).
- Perutz, M. F. Amyloid fibrils. Mutations make enzyme polymerize. *Nature* **385**, 773–775 (1997).
- Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Brünger, A. X-PLOR v3.1 Manual (Yale Univ. Press, New Haven, CT, 1992).
- Tronrud, D. E. Conjugate-direction minimization: an improved method for the refinement of macromolecules. *Acta Crystallogr. A* **48**, 912–916 (1992).
- Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. The MIDAS display system. *J. Mol. Graph.* **6**, 13–27 (1988).
- Lavery, R. & Sklenar, H. The definition of generalised helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63–91 (1988).
- Nicholls, A., Sharp, K. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Acknowledgements. We thank Y. Hunziker, L. Pellegrini and D. Sargent for their support and assistance, and A. Thompson at BM14 of the ESRF Grenoble for help during data collection. This research was supported in part by the Swiss National Science Foundation.

Correspondence and requests for materials should be addressed to T.J.R. The atomic coordinates have been deposited in the Brookhaven Protein Databank under code 1mmn.

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Extraordinary optical transmission through sub-wavelength hole arrays

T. W. Ebbesen^{*†}, H. J. Lezec[‡], H. F. Ghaemi^{*}, T. Thio^{*} & P. A. Wolff[§]

^{*} NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540, USA

[†] ISIS, Louis Pasteur University, 67000 Strasbourg, France

[‡] Micrion Europe GmbH, Kirchenstraße 2, 85622 Feldkirchen, Germany

[§] Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The desire to use and control photons in a manner analogous to the control of electrons in solids has inspired great interest in such topics as the localization of light, microcavity quantum electrodynamics and near-field optics^{1–6}. A fundamental constraint in manipulating light is the extremely low transmittivity of apertures smaller than the wavelength of the incident photon. While exploring the optical properties of submicrometre cylindrical cavities in metallic films, we have found that arrays of such holes display highly unusual zero-order transmission spectra (where the incident and detected light are collinear) at wavelengths larger than the array period, beyond which no diffraction occurs. In particular, sharp peaks in transmission are observed at wavelengths as large as ten times the diameter of the cylinders. At these maxima the transmission efficiency can exceed unity (when normalized to the area of the holes), which is orders of magnitude greater than predicted by standard aperture theory. Our experiments provide evidence that these unusual optical properties are due to the coupling of light with plasmons—electronic excitations—on the surface of the periodically patterned metal film. Measurements of transmission as a function of the incident light angle result in a photonic band diagram. These findings may find application in novel photonic devices.

A variety of two-dimensional arrays of cylindrical cavities in metallic films were prepared and analysed for this study. Typically, a silver film of thickness $t = 0.2 \mu\text{m}$ was first deposited by evaporation on a quartz substrate. Arrays of cylindrical holes were

fabricated through the film by sputtering using a Micrion focused-ion-beam (FIB) System 9500 (50 keV Ga ions, 5 nm nominal spot diameter). The individual hole diameter d was varied between 150 nm and $1 \mu\text{m}$ and the spacing between the holes (that is, the periodicity) a_0 , was between 0.6 and $1.8 \mu\text{m}$. The zero-order transmission spectra, where the incident and detected light are collinear, were recorded with a Cary 5 ultraviolet–near infrared spectrophotometer with an incoherent light source, but the arrays were also studied on an optical bench for transmission, diffraction and reflection properties using coherent sources.

Figure 1 shows a typical zero-order transmission spectrum for a square array of 150 nm holes with a period a_0 of $0.9 \mu\text{m}$ in a 200 nm thick Ag film. The spectrum shows a number of distinct features. At wavelength $\lambda = 326 \text{ nm}$ the narrow bullseye plasmon peak is observed which disappears as the film becomes thicker. The most remarkable part is the set of peaks which become gradually stronger at longer wavelengths, increasingly so even beyond the minimum at the periodicity a_0 . There is an additional minimum at $\lambda = a_0 \sqrt{\epsilon}$ corresponding to the metal–quartz interface (where ϵ is the

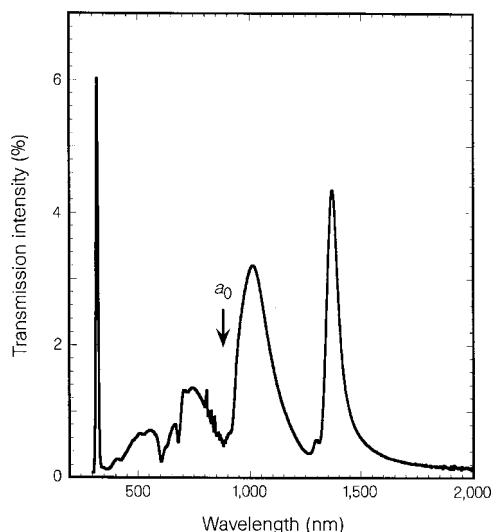


Figure 1 Zero-order transmission spectrum of an Ag array ($a_0 = 0.9 \mu\text{m}$, $d = 150 \text{ nm}$, $t = 200 \text{ nm}$).

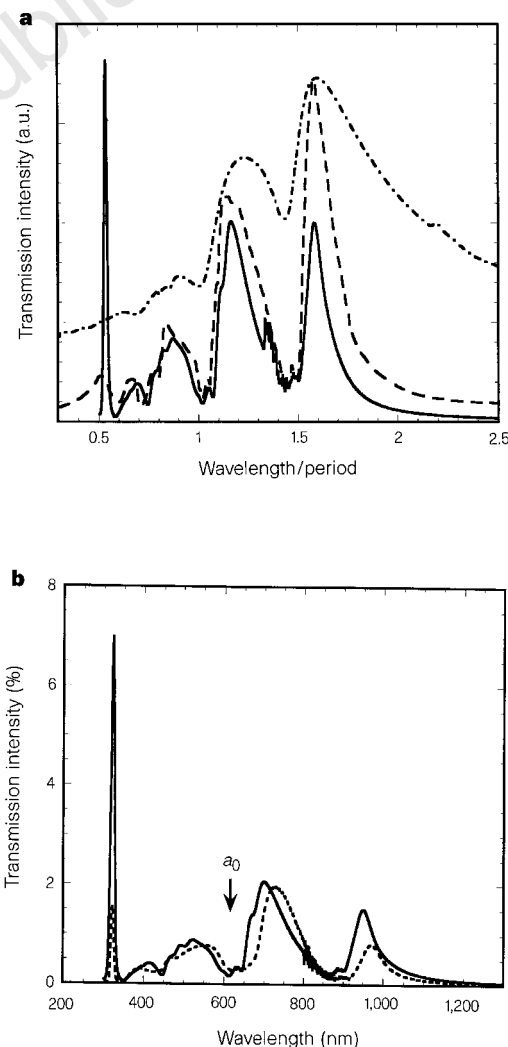


Figure 2 Effects of parameters on zero-order transmission spectra. **a**, Spectra for various square arrays as a function of λ/a_0 . Solid line: Ag, $a_0 = 0.6 \mu\text{m}$, $d = 150 \text{ nm}$, $t = 200 \text{ nm}$; dashed line: Au, $a_0 = 1.0 \mu\text{m}$, $d = 350 \text{ nm}$, $t = 300 \text{ nm}$; dashed-dotted line: Cr, $a_0 = 1.0 \mu\text{m}$, $d = 500 \text{ nm}$, $t = 100 \text{ nm}$. **b**, Spectra for two identical Ag arrays with different thicknesses. Solid line: $t = 200 \text{ nm}$; dashed line: $t = 500 \text{ nm}$ (this spectrum has been multiplied by 1.75 for comparison). For both arrays: $a_0 = 0.6 \mu\text{m}$; $d = 150 \text{ nm}$.

dielectric constant of the substrate). For $\lambda > a_0\sqrt{\epsilon}$, there is no diffraction from the array nor from the individual holes. As expected, the first-order diffraction spots can be seen to be grazing the surface (that is, the diffraction angle approaches 90°) as the wavelength approaches the period from below (this might, in fact, enhance the coupling to be discussed in the next paragraphs). The maximum transmitted intensity occurs at 1,370 nm, nearly ten times the diameter of an individual hole in the array. Even more surprising is that the absolute transmission efficiency, calculated by dividing the fraction of light transmitted by the fraction of surface area occupied by the holes, is ≥ 2 at the maxima. In other words, more than twice as much light is transmitted as impinges directly on the holes. Furthermore, the transmittivity of the array scales linearly with the surface area of the holes. This is all the more remarkable considering that the transmission efficiency of a single sub-wavelength aperture is predicted by Bethe⁷ to scale as $(r/\lambda)^4$ where r is the hole radius; accordingly for a hole of 150 nm diameter one expects a transmission efficiency on the order of 10^{-3} . In addition, the intensity (I) of the zero-order transmission from a grating is expected to decrease monotonically at larger wavelengths ($I \propto \lambda^{-1}$) (ref. 8). Therefore our results must imply that the array itself is an active element, not just a passive geometrical object in the path of the incident beam.

To understand the origin of this phenomenon, we tested the dependence on all the possible variables such as hole diameter, periodicity, thickness and type of metal. It is beyond the scope of this Letter to describe the details of all these experiments. Instead we summarize our observations and illustrate some of the key factors that determine the shape of these spectra. To begin with, the periodicity of the array determines the position of the peaks. The positions of the maxima scale exactly with the periodicity, as shown in Fig. 2a, independent of metal (Ag, Cr, Au), hole diameter and film thickness. The width of the peaks appears to be strongly dependent on the aspect ratio (t/d or depth divided by diameter) of the cylindrical holes (Fig. 2a). For $t/d = 0.2$, the peaks are very broad and just discernible and when the ratio reaches ~ 1 , the maximum sharpness is obtained. Further narrowing might depend on the

quality of the individual holes. The thickness dependence of the spectra is displayed in Fig. 2b for 0.2 and 0.5 μm Ag films. While the intensity of the bulk plasmon peak decreases rapidly in this range, that of the longer-wavelength peaks decreases approximately linearly with thickness. The spectra change significantly with the type of lattice, for example whether the array is a square or a triangular lattice.

Two important clues relating this phenomenon to surface plasmons (SPs) come from the following observations. One is the absence of enhanced transmission in hole arrays fabricated in Ge films which points to the importance of the metallic film. The other clue is the angular dependence of the spectra in metallic samples. The zero-order transmission spectra change in a marked way even for very small angles, as illustrated in Fig. 3 where the spectra were recorded every 2° . The peaks change in intensity and split into new peaks which move in opposite directions. This is exactly the behaviour observed when light couples with SPs in reflection gratings^{9–14}. SPs are oscillations of surface charges at the metal interface and are excited when their momentum matches the momentum of the incident photon and the grating as follows:

$$k_{\text{sp}} = k_x \pm nG_x \pm mG_y$$

where k_{sp} is the surface plasmon wavevector, $k_x = (2\pi/\lambda) \sin \theta$ is the component of the incident photon's wavevector in the plane of the grating and $G_x = G_y = 2\pi/a_0$ are the grating momentum wavevectors for a square array. Therefore if the angle of incidence θ is varied, the incident radiation excites different SP modes and by recording the peak energies as a function of k_x we obtain the dispersion relation shown in Fig. 4. This figure reveals the band structure of SP in the two-dimensional array and it clearly demonstrates the presence of gaps with energies around 30 to 50 meV; these are due to the lifting of the degeneracy by SPs interacting with the lattice. An extrapolation of the dispersion curves yields an intercept with the k -axis at a value which is within the experimental error ($\sim 10\%$) of that expected from the periodicity ($2\pi/a_0$). The results of Figs 3 and 4 are also sensitive to polarization as expected,

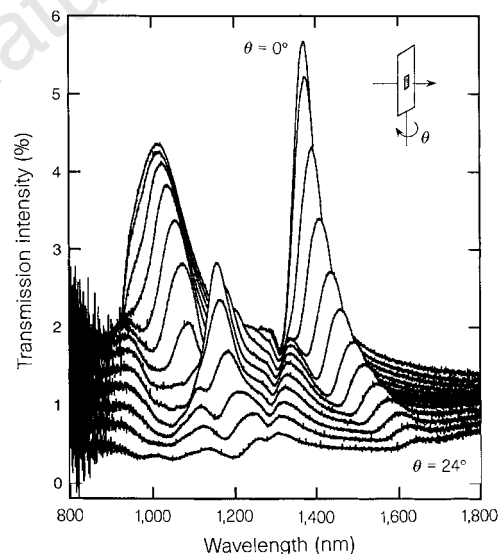


Figure 3 Zero-order transmission spectra as a function of incident angle of the light. Spectra were taken every 2° up to 24° for a square Ag array ($a_0 = 0.9 \mu\text{m}$, $d = 150 \text{ nm}$, $t = 200 \text{ nm}$). The individual spectra are offset vertically by 0.1% from one another for clarity.

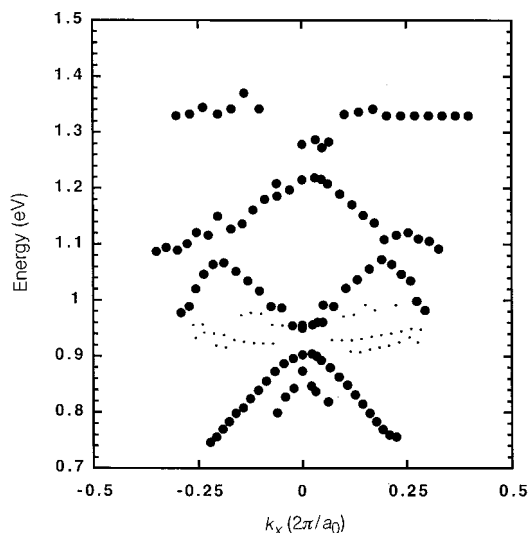


Figure 4 Dispersion curves (solid circles) along the $[10]$ direction of the array. These curves are extracted from the energy of the transmission peaks of Fig. 3. The momentum k_x is in the plane of the array and is given by $k_x = (2\pi/\lambda) \sin \theta$ where θ is the incident angle of the beam (the units are normalized to $2\pi/a_0$). The curves with the smaller dots correspond to peaks whose amplitudes are much weaker and may or may not be related to the band structure as they do not shift significantly with momentum.

but a clear assignment has not yet been possible because of inherent structure of the grains in the metal film.

In our experiments the coupling of light with SPs is observed in transmission rather than reflection, in contrast with previous work on SPs on reflection gratings. In those studies the coupling of light to SPs is observed as a redistribution of intensity between different diffracted orders. Even in transmission studies of wire gratings by Lochbihler¹⁵, the effect of SPs is observed through dips in zero-order transmitted light. There is an extensive literature on the infrared properties of wire grids which show a broad transmission centred at $\lambda \approx 1.2a_0$; this has been interpreted in terms of induction effects, in analogy with electric circuits^{16,17}. Our results demonstrate the strong enhancement of transmitted light due to coupling of the light with the SP of the two-dimensional array of sub-wavelength holes. Furthermore our results indicate a number of unique features that cannot be explained with existing theories. The SP modes on the metal–air interface are distinctly different from those at the metal–quartz interface. However, the spectra are identical regardless of whether the sample is illuminated from the metal or quartz side. At present we do not understand the detailed mechanism of the coupling between the SP on the front and back surfaces which results in larger than unity transmission efficiency of the holes. The thickness dependence (see Fig. 2b) suggests that the holes play an important part in mediating this coupling and that nonradiative SP modes are transferred to radiative modes by strong scattering in the holes.

In photonic bandgap arrays^{2,4}, the material is passive and translucent at all wavelengths except at the energies within the gap. In the present arrays, the material plays an active role (through the plasmons) and it is opaque at all wavelengths except those for which coupling occurs. The combination, or integration, of these two types of phenomena might lead to optical features that are very interesting from both fundamental and technological points of view. The demonstration of efficient light transmission through holes much smaller than the wavelength and beyond the inter-hole diffraction limit might, for example, inspire designs for novel near-field scanning optical microscopes⁶, or sub-wavelength photolithography. Theoretical analysis of the results would also be useful for gaining better insight into this extraordinary transmission phenomenon. Perhaps only then can we expect to grasp the full implications of these findings. □

Received 15 July; accepted 24 November 1997.

1. John, S. Localization of light. *Phys. Today* 32 (May 1991).
2. Yablonovitch, E. & Leung, K. M. Hope for photonic bandgaps. *Nature* 351, 278 (1991).
3. Dalichaouch, R., Armstrong, J. P., Schultz, S., Platzman, P. M. & McCall, S. L. Microwave localization by two-dimensional random scattering. *Nature* 354, 53–55 (1991).
4. J. D. Joannopoulos, Meade R. D. & Winn, J. N. *Photonic Crystals* (Princeton Univ. Press, Princeton, 1995).
5. Haroche, S. & Kleppner, D. Cavity quantum electrodynamics. *Phys. Today* 24 (January 1989).
6. Betzig, E. & Trautman, J. K. Near-field optics: Microscopy, spectroscopy, and surface modification beyond the diffraction limit. *Science* 257, 189–194 (1992).
7. Bethe, H. A. Theory of diffraction by small holes. *Phys. Rev.* 66, 163–182 (1944).
8. Born, M. & Wolf, E. *Principles of Optics* (Pergamon, Oxford, 1980).
9. Ritchie, R. H., Arakawa, E. T., Cowan, J. J. & Hamm, R. N. Surface-plasmon resonance effect in grating diffraction. *Phys. Rev. Lett.* 21, 1530–1533 (1968).
10. Raether, H. *Surface Plasmons* (Springer, Berlin, 1988).
11. Chen, Y. J., Koteles, E. S., Seymour, R. J., Sonek, G. J. & Ballantyne, J. M. Surface plasmons on gratings: coupling in the minigap regions. *Solid State Commun.* 46, 95–99 (1983).
12. Kitson, S. C., Barnes, W. L. & Sambles, J. R. Full photonic band gap for surface modes in the visible. *Phys. Rev. Lett.* 77, 2670–2673 (1996).
13. Watts, R. A., Harris, J. B., Hibbins, A. P., Preist, T. W. & Sambles, J. R. Optical excitations of surface plasmon polaritons on 90 and 60 bi-gratings. *J. Mod. Opt.* 43, 1351–1360 (1996).
14. Derrick, G. H., McPhedran, R. C., Maystre, D. & Neviere, M. Crossed gratings: a theory and its applications. *Appl. Phys.* 18, 39–52 (1979).
15. Lochbihler, H. Surface polaritons on gold-wire gratings. *Phys. Rev. B* 50, 4795–4801 (1994).
16. Ulrich, R. Far-infrared properties of metallic mesh and its complementary structure. *Infrared Phys.* 7, 37–55 (1967).
17. Larsen, T. A survey of the theory of wire grids. *I.R.E. Trans. Microwave Theory Techniques* 10, 191–201 (1962).

Acknowledgements. We thank S. Kishida, G. Bugmann and J. Giordmaine for their encouragement, and R. Linke, R. McDonald, M. Treacy, J. Chadi and C. Tsai for discussions. We also thank G. Lewen, G. Seidler, A. Krishnan, A. Schertel, A. Dziesiaty and H. Zimmermann for assistance.

Correspondence should be addressed to T.W.E.

Connecting atomistic and mesoscale simulations of crystal plasticity

Vasily Bulatov*, Farid F. Abraham†, Ladislav Kubin‡, Benoit Devincre‡ & Sidney Yip*

* Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† IBM Research Division, Almaden Research Center, San Jose, California 95120, USA

‡ Laboratoire d'Etude des Microstructures (CNRS-ONERA), 29 Av. de la Division Leclerc, BP72, 92322 Chatillon Cedex, France

A quantitative description of plastic deformation in crystalline solids requires a knowledge of how an assembly of dislocations—the defects responsible for crystal plasticity—evolves under stress¹. In this context, molecular-dynamics simulations have been used to elucidate interatomic processes on microscopic ($\sim 10^{-10}$ m) scales², whereas ‘dislocation-dynamics’ simulations have explored the long-range elastic interactions between dislocations on mesoscopic ($\sim 10^{-6}$ m) scales³. But a quantitative connection between interatomic processes and behaviour on mesoscopic scales has hitherto been lacking. Here we show how such a connection can be made using large-scale (100 million atoms) molecular-dynamics simulations to establish the local rules for mesoscopic simulations of interacting dislocations. In our molecular-dynamics simulations, we observe directly the formation and subsequent destruction of a junction (a Lomer–Cottrell lock) between two dislocations in the plastic zone near a crack tip: the formation of such junctions is an essential process in plastic deformation, as they act as an obstacle to dislocation motion. The force required to destroy this junction is then used to formulate the critical condition for junction destruction in a dislocation-dynamics simulation, the results of which compare well with previous deformation experiments⁴.

Recent molecular-dynamics (MD) studies of dynamic crack propagation using 100 million atoms have produced massive amounts of atomic-level details on the spontaneous emission of dislocation loops from the crack tip which result in shielding, blunting and crack arrest^{2,5}. A Lennard–Jones potential is used to simulate the response of a face-centred cubic (f.c.c.) structure characterized by a low stacking-fault energy. Figure 1 shows the evolution of a ‘flower’ of dislocation loops emitted from a crack tip advancing under the action of a transverse applied strain. We shall focus on the evolution of three particular dislocation loops shown in Fig. 1, labelled 1, 2 and 3. In this sequence one sees explicitly, for the first time to our knowledge, how a dislocation junction is first formed (Fig. 1b) and then partially destroyed in a cooperative process which involves the formation of a second junction while the first junction is being unzipped (Fig. 1c). The atomic configuration of the core of the first junction, known as a Lomer–Cottrell lock, is shown in Fig. 2. Dislocation configurations of this kind are believed to be one of the principal mechanisms of strain hardening in f.c.c. metals with low stacking-fault energy.

The process of dislocation junction formation which we have identified in the MD simulation has been discussed only in schematic terms in classical textbooks^{1,6,7}, but no explicit details produced by quantitative calculations in three dimensions have (to our knowledge) been given before. The rather intricate exchange between the two junctions, apparently unforeseen and unknown, is driven by the balance between the crack-tip stress, dislocation interaction forces, the line tension forces on the mobile dislocation segments, and the constraint forces from the sessile (immobile) junctions. The destruction of Lomer–Cottrell locks has been

thought to involve rather large stresses, because the junction segment is sessile and the whole configuration is fully extended. As shown by the above sequence (Fig. 1), junction destruction is likely to proceed via 'unzipping', that is, motion of the dislocation node along the junction line, as was anticipated by the Friedel–Saada model¹. If this were indeed the case, then the fate of a given junction should be governed primarily by the force driving nodal

displacement, while junction extension, as determined by the stacking-fault energy, should be of secondary importance. Another implication of the observed sequence is that a quantitative physical theory of strain hardening and dynamic recovery behaviour may need to include three-body and higher-order dislocation reactions.

The present example illustrates how large-scale atomistic simulations can provide new insights into possible mechanisms of close-

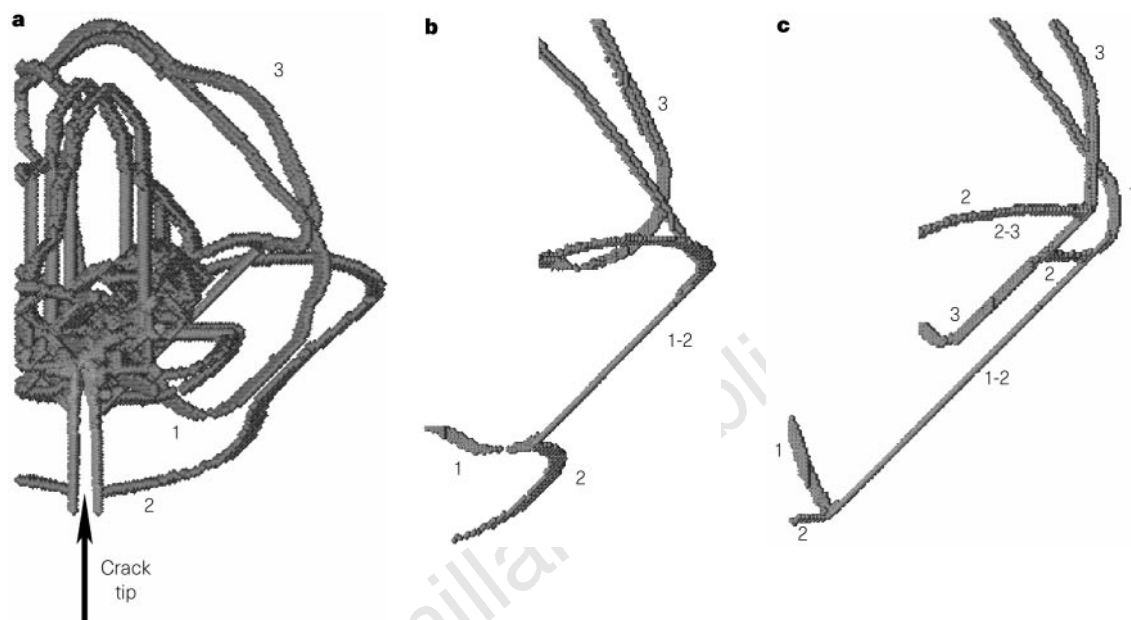


Figure 1 Evolution sequence of dislocation loops emitted spontaneously from an advancing crack tip which is suddenly arrested. In displaying the MD results, only atoms with energies exceeding the ideal bulk value by $\geq 3\%$ are shown. **a**, The expansion (spreading) of a number of dislocation loops, with loops 1 and 3 gliding in parallel (111) planes and loop 2 gliding in an intersecting (111) plane. As loops 1 and 2 approach each other, their leading segments are seen to attract each other and line up in parallel. These dislocations, on analysis, are found to be Shockley partial (imperfect) dislocations, with co-planar loops having the same Burgers vector, $a/6[11\bar{2}]$ where a is the lattice parameter, while that of loop 2 is $a/6[2\bar{1}1]$. In **b**, the leading segments of 1 and 2 have merged, forming a junction dislocation 1–2, or 'zipper', stretching along $[10\bar{1}]$ direction, with Burgers vector

equal to the vector sum of the Burgers vectors of the two parent dislocations: $a/6[11\bar{2}] + a/6[2\bar{1}1] = a/6[10\bar{1}]$. This reaction product, an imperfect dislocation known as a stair-rod, is sessile (immobile) because its slip plane, defined as the plane containing the line direction and Burgers vector, is not a glide plane in the f.c.c. structure. In **c**, loop 3 now glides in a plane parallel to loop 1. The elastic interaction between loop 3 and junction 1–2 is repulsive, causing loop 3 to line up along the junction. As a part of loop 3 touches the free end of the intersecting loop 2 (on the right), another junction 2–3 of the same Lomer–Cottrell type forms along the stretched segment of loop 3, while simultaneously unzipping the first junction 1–2. (Further details of fracture simulations are available at <http://www.almaden.ibm.com/st/Simulate/Fracture>)

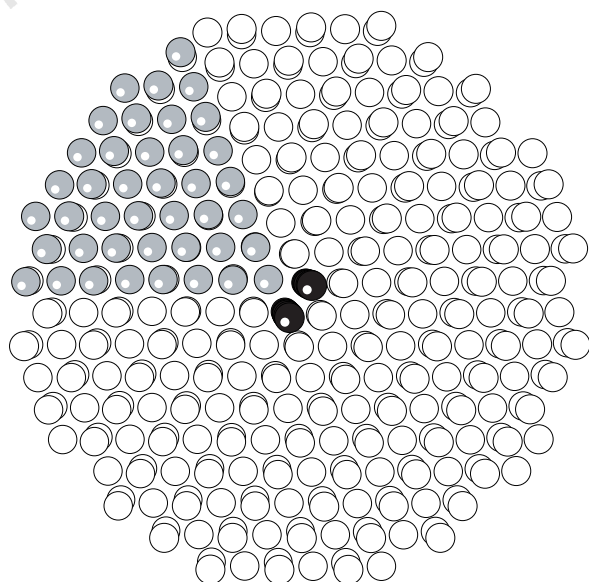


Figure 2 Atomic core structure of the junction dislocation viewed along $[10\bar{1}]$ junction line. The junction (black atoms) appears at the intersection of two stacking faults resulting from two intersecting partial dislocations, each of a mixed, 30° character with respect to the junction line direction. In terms of atomic displacement, the combination of the two partials is equivalent to carving out a triangular wedge (grey atoms) and displacing the wedge matter away from the junction by $a/6[10\bar{1}]$ and along the junction line by $a/4[10\bar{1}]$. The product dislocation is a sessile stair-rod whose core is narrow and energy is low, as indicated by the fact that only two rows of atoms (black) have energies in excess of the 3% threshold. This stair-rod is a part of the so-called Lomer–Cottrell lock⁶ which has been observed in experiments⁹.

range dislocation interactions. On the other hand, dislocation behaviour is also governed by long-range elastic interactions; a natural length scale for the collective response of dislocation microstructures is in the micrometre range. The appropriate method to treat the elastic interactions is a mesoscopic approach known as dislocation-dynamics. In this method, dislocations are represented by piecewise straight segments connecting the nodes of a regular three-dimensional mesh whose symmetries are those of the underlying crystal structure. Dislocation-dynamics simulations are currently used to examine collective dislocation behaviour on much larger scales, (typically $15\text{ }\mu\text{m}$) for f.c.c. crystals, and longer time intervals ($50\text{ }\mu\text{s}$) than is feasible with MD. Dislocation-dynamics studies have the potential to probe the macroscopic mechanical response of single crystals controlled by the collective behaviour of dislocations at high densities: such behaviours include the formation and evolution of dislocation microstructures under stress (Fig. 3) which ultimately determine the constitutive behaviour of single crystals in various stages of plastic deformation⁸. Of particular interest is the nature of the reciprocal scaling relationship between yield stress and the length scale of dislocation patterns, postulated to date as a universal empirical 'principle of similitude'.

In dislocation-dynamics simulations, computational efficiency is achieved through a less detailed description of dislocations in which

atomic degrees of freedom are replaced by piecewise straight lines, and a mesh spacing (a few nanometres) is used that is larger than the crystal lattice parameter. This means dislocation mobility and close-range interactions are not determined as atomic-level processes, but are specified by external parameters known as 'local rules'³. For this approach to be predictive, atomistic behaviour of dislocation cores has to be integrated into meso-scale dislocation-dynamics simulations. We therefore propose to establish a micro-to-meso connection in which the local rules are derived from the physically occurring dislocation core processes in an atomistic simulation. Figure 3c shows the formation, in a meso-scale simulation, of an attractive junction under conditions close to the MD simulation described above. The behaviour of this mesoscopic junction needs to be matched to the behaviour of its atomistic analogue, such as the one shown in Fig. 1b.

Evolution of the junction, zipping further or unzipping, will depend on the forces acting on the two mobile segments connected to each node at the junction's end (nodal segments). It is the junction strength that governs the strain-hardening properties of f.c.c. crystals (so-called stage II hardening). We now analyse the junction strength in terms of the minimum force required to drive the unzipping process. In general, the force F acting to separate the junction is the Peach–Koehler force, τb , where τ is the stress on the mobile nodal segment and b is the magnitude of its Burgers vector, times the effective arm-length of the segment l . The critical condition for unzipping is that F should be equal to the energy spent on separating a unit length of the junction and moving the resulting partials apart by a distance l . This energy is the sum of core W_c and elastic W_{el} contributions. Practically, for mesoscopic simulations, junction strength is treated through an effective junction stress $\tau_j = (W_c + W_{el})/bl$, which is then combined with the local stress τ acting on every segment connected to a junction node. Although τ_j is not a material parameter itself, as it depends on the length of the unzipping arm, the local rule for mobility of the nodal segments is nevertheless fully specified, once W_c and W_{el} are known. Whereas W_{el} can be calculated from elasticity theory, the core contribution W_c can only be determined atomistically.

We have carried out an explicit atomistic calculation to obtain the core contribution W_c . To avoid the complications of dealing with non-uniform stress distributions near a crack tip, we analyse the energetics of junction unzipping in the same crystal without the crack and under zero stress. The energy cost of junction unzipping was obtained by comparing the energy of a sufficiently large cylindrical slab ($559.6\text{ }\text{\AA}$ in diameter) centred around lock 1–2 (Fig. 1b) with the energy of the same slab for a configuration where partials 1 and 2 are lined up parallel at a distance d , similar to that shown in Fig. 1a. Presenting our result in a form scaled to the length l of the mobile segment, the effective junction stress is given by the above equation with $W_c(b) = -0.0542\text{ eV }\text{\AA}^{-1}$ and $W_{el}(l/b) = 0.106\ln(l/b)\text{ eV }\text{\AA}^{-1}$, $b = 1.65\text{ }\text{\AA}$. These numerical values were obtained for Lennard–Jones potential with well depth $\epsilon = 0.15\text{ eV}$ and length parameter $\sigma = 2.556\text{ }\text{\AA}$, nominally representing Cu. For the arm length of $0.5\text{ }\mu\text{m}$, the effective junction stress is 15.5 MPa , which compares very well with the junction strength parameter previously obtained by fitting mesoscopic simulation data to deformation experiments⁴.

The overlap of atomistic and continuum length scales, demonstrated in this work for the case of dislocation junctions, points to the feasibility of a critical connection between two different modelling methodologies. It should lead to significantly improved understanding of dislocation behaviour which up to now has been investigated mostly within the framework of linear elasticity. As well as the properties of dislocation locks and junctions, a host of other essential mechanisms could be elucidated, including the cross-slip properties of screw dislocations, the dynamics of the formation and motion of kink pairs, and the annihilation properties of closely spaced edge-dislocation dipoles. □

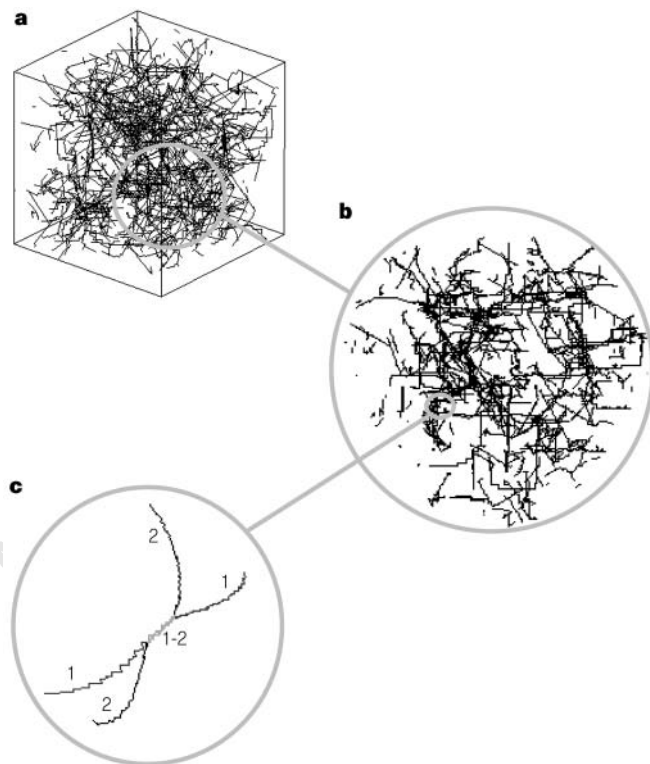


Figure 3 A composite of dislocation-dynamics simulation results. **a**, Snapshot of a dislocation-dynamics simulation where dislocations, at an initial density $2 \times 10^8\text{ cm}^{-2}$ and placed in the simulation box at random, move in response to external and interaction stresses. **b**, A magnified interior cross-section of thickness $1\text{ }\mu\text{m}$ extracted from the simulation box of **a**; it can be seen that the dislocations are beginning to organize into a cell pattern. **c**, Further magnification showing two dislocation lines 1 and 2 tied together by a junction 1–2. The junction, formed by two perfect dislocations with Burgers vectors $a/2[101]$ and $a/2[110]$ moving under applied stress on two intersecting slip planes, (111) and ($\bar{1}11$), respectively, is a perfect $a/2[011]$ dislocation. The parent dislocations mutually knit with each other until an equilibrium configuration is reached. The junction formation is mimicked by locking conditionally the knitted portions. (Further details of dislocation-dynamics simulations are available at <http://zig.onera.fr/lem/DisGallery>)

Received 1 October; accepted 30 December 1997.

1. Friedel, J. *Dislocations* (Pergamon, Oxford, 1964).
2. Abraham, F. F. *et al.* Instability dynamics in three-dimensional fracture: an atomistic approach. *J. Mech. Phys. Solids* **45**, 1461–1471 (1997).
3. Devincere, B. & Kubin, L. P. Mesoscopic simulations of dislocations and plasticity *Mater. Sci. Eng. A234*–236, 8–14 (1997).
4. Devincere, B. & Kubin, L. P. Simulations of forest interactions and strain hardening in FCC crystals. *Modelling Simul. Mater. Sci. Eng.* **2**, 559–570 (1994).
5. Abraham, F. F. On the transition from brittle to plastic failure in breaking a nanocrystal under tension. *Europhys. Lett.* **38**, 103–106 (1997).
6. Hirth, J. P. & Lothe, J. *Theory of Dislocations* 2nd edn 792–806 (Wiley, New York, 1982).
7. Nabarro, F. R. N. *Theory of Crystal Dislocations* (Oxford Univ. Press, London, 1967).
8. Prinz, F. & Argon, A. S. Dislocation cell formation during plastic deformation of copper single crystals. *Phys. Status Solidi A* **57**, 741–753 (1980).
9. Korner, A. & Karthaler, H. P. Weak-beam studies of composite dislocations gliding on (001) planes in Si. *Phil. Mag.* **44**, 275–284 (1981).

Acknowledgements. This work originated at a collaborative workshop on materials modelling held at the Institute for Theoretical Physics, UCSB, during March 1997. F.F.A. acknowledges the use of the Cornell Theory Center, which receives funding from the NSF and New York State; B.D. and L.K. acknowledge the support of CNRS through the GDR programme on mesoscopic simulations and modelling in metallurgy; V.B. and S.Y. acknowledge support from LLNL; and V.B. acknowledges support from the Alcoa Foundation.

Correspondence and requests for materials should be addressed to V.B. (e-mail: vasily@mit.edu).

Impact energy measurement in time-of-flight mass spectrometry with cryogenic microcalorimeters

G. C. Hilton*, John M. Martinis*, D. A. Wollman*, K. D. Irwin*, L. L. Dulcie*, Daniel Gerber†, Patrick M. Gillevet‡ & Damian Twerenbold†

* National Institute of Standards and Technology, Boulder, Colorado 80303, USA

† Institut de Physique de l'Université Rue A.-L. Breguet 1, CH-2000 Neuchâtel, Switzerland, and GenSpec SA, case postale 120, CH-2017 Boudry, Switzerland

‡ Institute of Bioscience, Bioinformatics, and Biotechnology, George Mason University, Manassas, Virginia 22210, USA

Time-of-flight mass spectrometry—most notably matrix-assisted laser-desorption-ionization time-of-flight (MALDI-TOF) spectrometry¹—is an important class of techniques for the study of proteins and other biomolecules². Although these techniques provide excellent performance for masses up to about 20,000 daltons, there has been limited success in achieving good mass resolution at higher masses. This is because the sensitivity of the microchannel plate (MCP) detectors used in most systems decreases rapidly with increasing particle mass, limiting the utility of MCP detectors for very large masses. It has recently been proposed that cryogenic particle detectors may provide a solution to these difficulties³. Cryogenic detectors measure the thermal energy deposited by the particle impact, and thus have a sensitivity that is largely independent of particle mass. Recent experiments^{4–6} have demonstrated the sensitivity of cryogenic particle detectors to single biomolecules, a quantum efficiency several orders of magnitude larger than the MCP detectors, and sensitivity to masses as large as 750,000 daltons. Here we present results demonstrating an order of magnitude better energy resolution than previous measurements, allowing direct determination of particle charge state during acceleration⁷. Although application of these detectors to practical mass spectrometry will require further development of the detectors and cryogenics, these detectors can be used to elucidate the performance-limiting processes that occur in such systems.

In the past several years, significant progress has been made in developing high-performance cryogenic detectors for applications such as infrared bolometry and X-ray, visible and ultraviolet spectroscopy. These detectors fall into two classes: non-equilibrium

devices such as superconducting tunnel junctions (STJs)^{8,9}, and equilibrium devices such as microcalorimeters^{10–13}. The previous experiments^{4,5} with cryogenic particle detectors for mass spectrometry used STJs, which fail to record the total impact energy because they are insensitive to the fraction of impact energy deposited as phonons with energy less than the superconducting gap. In this experiment we used a normal-insulator-superconductor microcalorimeter fabricated on a thin Si₃N₄ membrane, which detects all of the thermal energy deposited by a particle impact. Although the speed and energy resolution performance of our detector is sufficient for this preliminary experiment, experience with other applications indicate that significantly better performance should be possible in the future. It is important to note that the collection area of the cryogenic detectors studied is more than 10⁴ times smaller than typical MCP detectors. Although this size can be improved in the future, the overall sensitivity (the product of quantum efficiency and collection area) of this detector is significantly less than MCP detectors except at extremely large (>100,000 daltons) masses⁶.

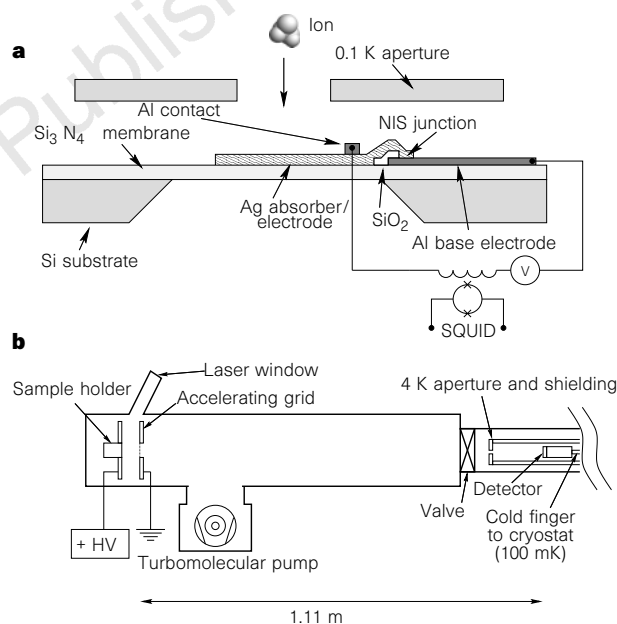


Figure 1 Details of experimental apparatus. **a**, Cross-section of the normal-insulator-superconductor (NIS) microcalorimeter. An accelerated molecule flies through a 200-μm-diameter 0.1 K aperture and strikes the absorber (a 100-nm-thick Ag film measuring 200 μm × 200 μm deposited on a 0.5-μm-thick Si₃N₄ membrane), raising the absorber temperature. This temperature rise is detected as a current pulse produced by a voltage-biased NIS junction, which consists of a thinly oxidized Al electrode in contact with the absorber. The junction current is measured by a SQUID preamplifier. Contacts to the detector are provided by superconducting Al leads which carry electrical current but do not conduct heat. **b**, Schematic view of our MALDI-TOF spectrometer indicating detector placement and infrared shielding. The infrared shielding consists of a 1-mm-diameter aperture cooled to 4 K and placed 10 cm away from the detector. Ions are created when laser light pulses (337 nm, 3 ns, maximum power 100 μJ) strike the probe plate. Ions are accelerated by an electric field between the probe plate and the grounded fine-mesh accelerating grid. Vacuum in the spectrometer is provided by a 110 l s⁻¹ turbomolecular pump, attaining a base pressure in the 100-mm-diameter flight tube of 2 × 10⁻⁵ Pa (1.5 × 10⁻⁷ torr). Two proteins were studied, lysozyme (mass 14,300 daltons) and bovine serum albumin (BSA, mass 66,430 daltons). In both cases proteins solutions were prepared by combining 1 mg of protein with 1 ml of a 0.1% trifluoroacetic acid H₂O solution. The matrix solution consisted of 100 mg sinapinic acid (mass 224 daltons) dissolved in 6 ml of ethanol and 4 ml H₂O. Probes were prepared by mixing 40 μl each of the probe and matrix solutions and allowing the mixture to dry in air on the probe plate.

Our detector, similar to that described elsewhere¹², is shown in cross-section in Fig. 1a, along with the external circuit used to bias and read out the detector. The detector current is measured using a high-speed low-noise series-array superconducting quantum interference device (SQUID) amplifier¹⁴. The detector is cooled to its operating temperature of 100 mK using a liquid-helium cryostat with an adiabatic demagnetization refrigerator. The cryostat is coupled to a MALDI-TOF spectrometer as shown in Fig. 1b.

The detector energy scale was calibrated using X-rays from a ⁵⁵Fe radioactive source mounted in the detector line of sight. The energy resolution of the detector was measured for X-rays using several thousand digitized waveforms. The energy of each event was determined using Weiner optimal filtering, and the resolution was determined by fitting the resulting energy histogram to a normal distribution. For X-rays we obtained an energy resolution of 92 eV full-width at half-maximum (FWHM). In Fig. 2a we show a digitized detector waveform obtained from one laser strike on a protein target of bovine serum albumin (BSA, mass 66,430 daltons).

In Fig. 2b we show a scatter plot of impact energies versus arrival times for a sample of BSA in a sinapinic acid matrix (mass 224 daltons) for an accelerating voltage, U , of 20 kV. The plot shows a clear energy banding, caused by the discrete ionization states of the particles. The utility of impact energy resolution is clearly shown by examining two groupings of points, one at $t_1 = 146 \mu\text{s}$ and $E_1 = 10 \text{ keV}$, and the other at $t_2 = 103 \mu\text{s}$ and $E_2 = 20 \text{ keV}$. Calculating the mass corresponding to t_1 using $m = (2zUt^2)/l^2$ where l is the length of the flight tube and taking charge $z = e$, we find that $m_1 = 66,600$ daltons. If we similarly calculate the mass correspond-

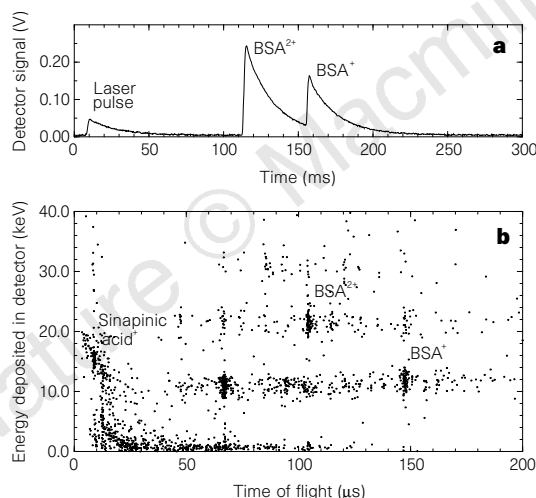


Figure 2 Experimental data obtained for the protein BSA. **a**, An example time trace for one laser event with three detector pulses. The first pulse is due to reflected laser light striking the detector, the remaining pulses are due to ions and are identified as BSA^{2+} and BSA^+ . **b**, A scatter plot of detector energy versus ion time of flight for BSA with an acceleration voltage of 20 kV. Clear energy banding is shown for ions of different charge. The ratio of energy deposited in the detector to ion kinetic energy (E_d/zU) is 0.59 for BSA^+ and 0.74 for singly charged sinapinic acid. This scatter plot was generated by extracting time and energy information from time traces for several thousand laser events. For each laser pulse we acquired a time trace consisting of 4,096 samples of the SQUID preamplifier output at either 50-ns or 100-ns intervals using a 20 MHz 12-bit analog-to-digital converter. Particle impact times were determined by an edge detection algorithm applied to the time traces and particle impact energies were extracted by comparing the impact waveforms with a calibrated waveform derived from X-ray events. The signal rise time ($1.2 \mu\text{s}$) allowed us to determine the time of impact within $\sim 200 \text{ ns}$. Potential difficulties of pulse pile-up due to the slow detector fall time ($17 \mu\text{s}$) are mitigated by the detector linearity, the 12 bit digitization, and the small number of particle strikes per launch event.

ing to t_2 with $z = 2e$, we find $m_2 = 66,400$ daltons. Thus we associate the first grouping with BSA^+ and the second grouping with BSA^{2+} . Additional groupings in Fig. 2b include sinapinic acid ($m = 224$, $t = 10 \mu\text{s}$) and an unidentified fragment with $m \approx 14,000$ daltons ($t = 67 \mu\text{s}$). There is also a small grouping at the same time-of-flight as the BSA^+ group with twice the energy of the BSA^+ group. This group is due to either doubly charged BSA dimers or the simultaneous arrival of two singly charged BSA monomers.

A surprising feature of Fig. 2b is that the thermal energy E_d deposited in the microcalorimeter by a particle impact is roughly half of the particle's kinetic energy zU . This difference implies that a significant fraction of the particle's kinetic energy is not converted to thermal energy in the detector. Thus impact cannot be modelled as a rigid molecule striking, and sticking to, the detector. Processes that might account for the missing energy include fragmentation of the molecule hitting the detector and ejection of molecular fragments.

To investigate further the dependence of impact energy on molecule type, we show in Fig. 3 the ratio of deposited to kinetic energy versus kinetic energy for three different molecules. At low

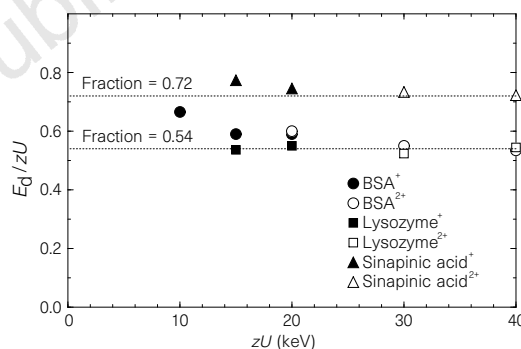


Figure 3 Plot of the ratio of impact energy to kinetic energy versus kinetic energy for each ion type. The ratio is determined by dividing the centroid of the impact energy distribution for an identified particle grouping by the kinetic energy of that grouping. For the heavier ions (lysozyme, BSA) this fraction approaches 0.54 for large acceleration potentials, whereas for the lighter ion (sinapinic acid) the asymptote is 0.72.

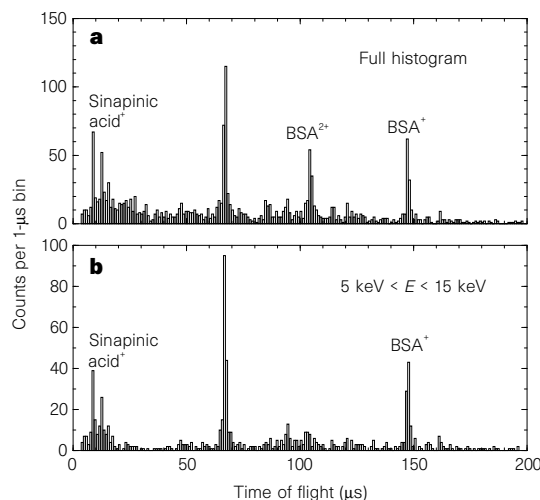


Figure 4 Time-of-flight spectrum for BSA with an acceleration voltage of 20 kV. **a**, Full histogram with all acquired points. **b**, Histogram containing only those points whose deposited energy is between 5 and 15 keV. The peak due to doubly ionized BSA, as well as much of the background, is removed by this selection.

kinetic energies, we see indications of an increase in fractional energy deposition, as might be expected as there is less energy available to deform the ion. Presumably E_d/zU approaches 1 at very low kinetic energies. At large kinetic energies, the proteins deposit about half of their kinetic energy in the detector, while the much lighter sinapinic acid deposits nearly three-quarters of its kinetic energy. Because there are many more bonds in the larger molecules, it is not surprising that more of the kinetic energy is lost in impact ejection or fragmentation. Although we do not fully understand the origin of the asymptotic 50% impact energy observed at high kinetic energies for lysozyme and BSA, preliminary experiments with immunoglobulin G (IgG, 155,000 daltons) show that this same asymptotic limit holds for larger masses. Because the total bond energy of sinapinic acid is less than the impact energy deficit, we believe that this is evidence that ejection is the dominant energy loss mechanism. In the only previous measurement of fractional energy deposition using STJs⁵, the measured impact energy was ~20% of the kinetic energy. The lower value of E_d/zU obtained with the STJ detectors is not surprising, given the phonon loss mechanisms previously discussed.

Histograms of arrival times for BSA are plotted in Fig. 4a using all particle impacts, and in Fig. 4b using only those events with impact energies between 5 and 15 keV. In the selective plot of Fig. 4b, both the background and the BSA²⁺ peak have been greatly reduced. Similar histograms were used to determine the mass and energy resolution of the mass spectrometer. For the BSA monomer peak, we obtain a timing resolution of 1.3 μ s (FWHM), implying a mass resolution $M/\Delta M$ of 56, similar to results obtained elsewhere¹⁵. An energy resolution of 1.7 keV (FWHM) was obtained, which is

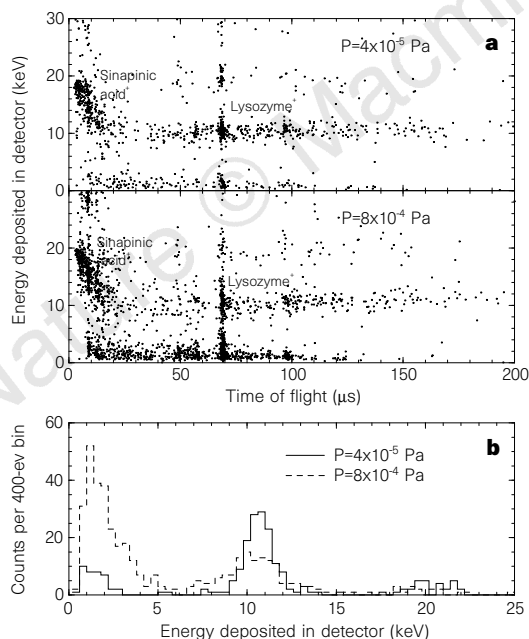


Figure 5 Experimental data showing additional scattering and changes in impact energy resolution due to gas phase scattering. **a**, Scatter plots for lysozyme (mass 14,300 daltons) taken at two different pressures. At the higher pressure, many more ions with low impact energy are observed. This effect is due to gas-phase scattering and fragmentation in the flight tube. The mean free path for lysozyme collision with ambient gas molecules is 1/3 of the flight tube length in the high-pressure case and 6 times the flight tube length in the low-pressure case. **b**, Histogram of impact energies for the expected arrival time of lysozyme⁺ ($68 \mu s < t < 70 \mu s$). The data at higher pressure show both a low-energy peak and an enhanced low-energy tail of the 11-keV peak. These features are due to small and large fragments of the lysozyme caused by gas-phase fragmentation in the free-flight region.

significantly worse than both the 92 eV (FWHM) resolution obtained with X-rays and the kinetic energy uncertainty implied by the time-of-flight uncertainty, inferring that the energy lost to ejection and fragmentation during impact must be variable.

Measurement of impact energy with the microcalorimeter allows us to explore the effect of background gas pressure in the spectrometer. As background pressure increases, gas-phase collisions in both the acceleration and free-flight regions increase, decreasing the quality of the spectra. The effect of pressure is clearly shown in Fig. 5a, which shows scatter-plot spectra for lysozyme (mass 14,300 daltons) taken at 4×10^{-5} Pa (3×10^{-7} torr) and 8×10^{-4} Pa (6×10^{-6} torr). At high pressure, we observe a significant increase in the number of low-energy (<5 keV) particles arriving at the predicted time of flight for lysozyme. We attribute this effect to fragmentation of the lysozyme ions owing to collisions with background gas in the free-flight region of the spectrometer. Fragments produced during free flight carry less kinetic energy, but travel at velocities nearly equal to that of the original molecule and arrive at the expected time for lysozyme. On the other hand, fragmentation during acceleration would affect the velocity and cannot explain these events. In Fig. 5b we show the effect on this fragmentation on impact energy. Here we plot a histogram of impact energies for those particles arriving with near the predicted time of flight for lysozyme. The histograms show that at low pressure there are many more low-energy events which are caused by small fragments, and a broadening of the primary 11-keV peak by a low-energy tail which is caused by larger fragments.

Insight into the effect of gas phase collisions can be gained by estimating the cross-sections and mean free paths of the proteins. From X-ray structure data for lysozyme¹⁶, we estimate the molecular cross-section for gas-phase scattering to be 15 nm². With this cross-section, the mean free path for lysozyme at a nitrogen gas pressure of 4×10^{-5} Pa is ~6 m, whereas at 8×10^{-4} Pa the mean free path is 0.3 m. Thus at the higher pressure considered in Fig. 5, lysozyme will typically undergo 3–4 collisions in traversing the flight tube. Assuming hard-sphere elastic collisions between a 20-keV lysozyme ion and an ambient-temperature nitrogen molecule, the maximum energy transferred per collision is 160 eV. This energy is more than sufficient to break bonds in lysozyme, supporting our conclusion that fragmentation occurs during flight. These results are similar to those reported elsewhere¹⁷, where fragmentation caused in the free-flight region was measured by subsequent deceleration of ions, allowing measurement of the relative concentration of neutral fragments relative to unfragmented ions to yield the fragmentation cross-section. For a protein similar to lysozyme (cytochrome c), a cross-section of 14.1 nm² was measured, consistent with our results.

A similar calculation of the cross-section for BSA shows that gas-phase collisions cannot explain the mass resolution obtained with our spectrometer. Using X-ray data¹⁸ we estimate the cross-section of BSA to be 30 nm², giving a mean free path of 3 m at the normal operating pressure. Assuming that the maximum energy transferred per gas collision is 80 eV, and that each molecule undergoes on average less than one collision, the worst case estimate for mass resolution $M/\Delta M$ is 250, which is better than the observed value of 56. In the present experiments, where mass resolution is not limited by gas collisions or detector response time, we conclude that the greatest contributions to mass uncertainty arise during launch and acceleration. □

Received 17 April; accepted 16 December 1997.

- Karas, M. & Hillenkamp, F. Laser desorption/ionization of proteins and molecular masses exceeding 10000 Daltons. *Anal. Chem.* **188**, 2299–2301 (1988).
- Siuzdak, G. The emergence of mass spectrometry in biochemical research. *Proc. Natl. Acad. Sci. USA* **91**, 11290–11297 (1994).
- Twerenbold, D. Biopolymer mass spectrometer with cryogenic particle detectors. *Nucl. Instrum. Meth. Phys. Res. A* **370**, 253–255 (1996); US Patent No. 5,640,010 (1997).
- Twerenbold, D. et al. Detection of single macromolecules using a cryogenic particle detector coupled to a biopolymer mass spectrometer. *Appl. Phys. Lett.* **68**, 3503–3505 (1996).
- Frank, M., Mears, C. A., Labov, S. E. & Benner, W. H. High-efficiency detection of 66 000 Da protein

- molecules using a cryogenic detector in a matrix-assisted laser desorption/ionization time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **10**, 1946 (1996).
6. Labov, S. E. *et al.* Cryogenic detector development at LLNL: ultraviolet, X-ray, gamma-ray and biomolecule spectroscopy. In *Proc. 7th Int. Workshop on Low Temperature Detectors* (ed. Cooper, S.) 82–95 (Max Planck Institute of Physics, Munich, 1997).
 7. Benner, W. H. *et al.* Simultaneous measurement of flight time and energy of large matrix-assisted laser desorption ions with a superconducting tunnel junction detector. *J. Am. Soc. Mass Spectrom.* **8**, 1094–1102 (1997).
 8. Mears, C. A. *et al.* Analysis of pulse shape from a high resolution superconducting tunnel junction X-ray spectrometer. *Nucl. Instrum. Meth. Phys. Res. A* **370**, 53–56 (1996).
 9. Peacock, A. *et al.* Single optical photon detection with a superconducting tunnel junction. *Nature* **381**, 135–137 (1996).
 10. McCammon, D., Cui, W., Juda, M., Morgenthaler, J. & Zhang, J. Thermal calorimeters for high resolution X-ray spectroscopy. *Nucl. Instrum. Meth. Phys. Res. A* **326**, 157–165 (1993).
 11. Irwin, K. D., Hilton, G. C., Wollman, D. A. & Martinis, J. M. X-ray detection using a superconducting transition-edge sensor microcalorimeter with electrothermal feedback. *Appl. Phys. Lett.* **69**, 1945–1947 (1996).
 12. Nahum, M. & Martinis, J. M. Hot-electron microcalorimeters as high-resolution x-ray detectors. *Appl. Phys. Lett.* **66**, 3203–3205 (1995).
 13. Martinis, J. M. Hot-electron-microcalorimeters with 0.25 mm² area. *Nucl. Instrum. Meth. Phys. Res. A* **370**, 171–172 (1996).
 14. Welty, R. P. & Martinis, J. M. A series array of DC SQUIDS. *IEEE Trans. Mag.* **27**, 2924–2926 (1991).
 15. Schreiner, M., Strupat, K., Lottspeich, F. & Eckershorn, C. Ultraviolet matrix assisted laser desorption ionization-mass spectrometry of electroblooded proteins. *Electrophoresis* **17**, 954–961 (1996).
 16. Vaney, M. C., Maignan, S., Ries-Kautt, M. & Ducruix, A. *Protein Data Bank File No. 1931* (Brookhaven National Laboratory Protein Data Bank, 1995).
 17. Spengler, B., Kirsch, D. & Kaufmann, R. Fundamental aspects of postsources decay in matrix-assisted laser desorption mass spectrometry. 1. Residual gas effects. *J. Phys. Chem.* **96**, 9678–9684 (1992).
 18. He, X. M. & Carter, D. C. Atomic structure and chemistry of human serum albumin. *Nature* **358**, 209–215 (1992).

Correspondence and requests for materials should be addressed to G.H. (e-mail hilton@boulder.nist.gov).

Increased stratospheric ozone depletion due to mountain-induced atmospheric waves

K. S. Carslaw*, M. Wirth†, A. Tsias*, B. P. Luo*, A. Dörnbrack†, M. Leutbecher†, H. Volkert†, W. Renger†, J. T. Bacmeister‡, E. Reimer§ & Th. Peter*

* Max-Planck-Institut für Chemie, Postfach 3060, Mainz 55020, Germany

† DLR, Oberpfaffenhofen, 32230 Wessling, Germany

‡ Code 7641, Naval Research Laboratory, Washington DC, 20375, USA

§ Institut für Meteorologie, Carl-Heinrich-Becker Weg, 12165, Berlin 41, Germany

Chemical reactions on polar stratospheric cloud (PSC) particles are responsible for the production of reactive chlorine species (chlorine 'activation') which cause ozone destruction¹. Gas-phase deactivation of these chlorine species can take several weeks in the Arctic winter stratosphere, so that ozone destruction can be sustained even in air parcels that encounter PSCs only intermittently^{2,3}. Chlorine activation during a PSC encounter proceeds much faster at low temperatures when cloud particle surface area and heterogeneous reaction rates are higher⁴. Although mountain-induced atmospheric gravity waves are known to cause local reductions in stratospheric temperature of as much as 10–15 K (refs 5–9), and are often associated with mesoscale PSCs^{10–12}, their effect on chlorine activation and ozone depletion has not been considered. Here we describe aircraft observations of mountain-wave-induced mesoscale PSCs in which temperatures were 12 K lower than expected synoptically. Model calculations show that despite their localized nature, these PSCs can cause almost complete conversion of inactive chlorine species to ozone-destroying forms in air flowing through the clouds. Using a global mountain-wave model⁸, we identify regions where mountain waves can develop, and show that they can cause frequent chlorine activation of air in the Arctic stratosphere. Such mesoscale processes offer a possible explanation for the under-prediction of reactive chlorine concentrations and ozone depletion

rates calculated by three-dimensional models of the Arctic stratosphere^{13–17}.

A stratospheric mountain-wave-induced PSC observed by airborne lidar over Scandinavia is shown in Fig. 1a. An ice cloud ('mother-of-pearl cloud') is apparent in the lidar signal as the yellow and red streaks of high aerosol backscatter ratio at around 22 km altitude. The presence of this cloud indicates strong localized cooling due to mountain-wave activity because the ice frost point at this altitude ($T_{\text{frost}} \approx 185.7$ K for 5 parts per million by volume, p.p.m.v., water vapour) is considerably lower than the temperature of 193 K derived from synoptic analyses from the European Centre for Medium-Range Weather Forecasts (ECMWF). The ice cloud is surrounded by a liquid PSC, which is visible in the lidar signal as a distinct wave-disturbed aerosol layer between 21 and 24 km altitude¹². This probably formed because of absorption of HNO₃ and H₂O by the background liquid H₂SO₄/H₂O aerosols¹⁸. On the day of measurement, the research aircraft made several flights across and parallel to the mountains¹². All the measurements are consistent with a mountain-induced PSC extending over several hundred kilometres, with dense ice clouds over the highest orography. Analysis of other nearby clouds and of the general mesoscale situation¹² indicates that the cloud in Fig. 1, including the liquid PSC, was probably embedded in a broader mesoscale minimum temperature.

Here we use data from this short flight segment to demonstrate the potential effect of even such small mesoscale PSCs on stratospheric chemistry. Two features of this flight make it ideal for studying the effect of such clouds on the chemistry of air parcels flowing over the mountains. First, the observation was made by flying directly upwind, which we term quasi-lagrangian. Second, the wave-cloud can be considered as quasi-stationary during the short time required for air parcels to traverse the wave, which is ~30 minutes. This was confirmed by repeated measurements of other mountain-wave clouds on the same day which revealed quasi-stationarity on a timescale of ~4 hours. Also, simulations using the mesoscale model MM5¹² indicate that gravity waves in this region were persistent for at least 1 hour. Together, the quasi-lagrangian observation and the stationarity of the cloud allow the particle streaks to be considered as air-parcel trajectories. One such trajectory is indicated by the white line in Fig. 1a, with a temperature shown in Fig. 1c. Thus, the trajectory that we use to study air-parcel microphysics and chemistry was constructed from aerosol layers apparent in the lidar signal, rather than from predictions of the MM5 mesoscale model. Although the MM5 model reproduces very well the general pattern of cooling over the mountains on the day of measurement¹², resolution limitations mean that the small-scale structure and deep temperature minima apparent in Fig. 1a cannot easily be resolved.

The air-parcel temperature is critical for the microphysical and chemical modelling of the cloud. The existence of ice clearly indicates a departure from synoptic conditions, but the minimum temperature attained in the wave is even lower than the frost point. This minimum temperature can be calculated as follows: at the point of maximum backscatter (point 1 in Fig. 1a) the ice particles are transiently in equilibrium with the gas phase, which implies that the temperature must be below the frost point ($T < T_{\text{frost}}$). How much point 1 is below the frost point depends on the amount of water vapour condensed as ice. From our microphysical model simulation, we calculate the actual temperature at point 1 to be 183.8 ± 0.5 K. Temperatures at other locations in the wave can then be calculated from this reference temperature by assuming dry adiabatic behaviour. The minimum temperature in the wave occurs ~280 m higher in altitude (point 2 in Fig. 1a) and is therefore ~181 K, some 12 K lower than synoptic. The rather constant offset of 8–10 K between the synoptic temperature and the temperature in the wave reflects the fact that the PSC is embedded in a larger mesoscale minimum, and does not indicate a significant warm bias

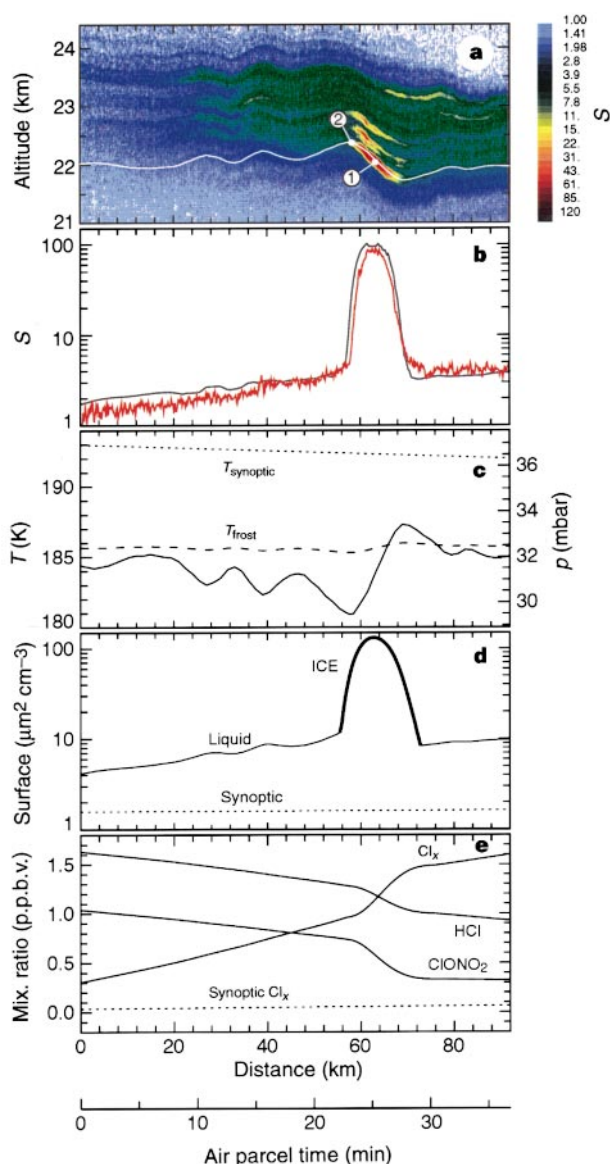
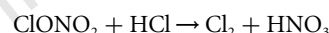


Figure 1 Lidar measurements of a mountain-wave PSC. These were made on 15 January 1995 between 67.2°N, 17.2°E and 67.5°N, 19.2°E from a Transall research aircraft. Also shown is the effect of the cloud on particle growth and chemistry. The time axis refers to an air parcel flowing through the cloud with a wind speed of 45 m s⁻¹, as determined from MM5 mesoscale model calculations¹². **a**, Backscatter ratio *S* at 532-nm laser wavelength. **b**, Calculated *S* (black line) and observed *S* (red line) along the trajectory indicated by the white line in **a**. In terms of the backscatter coefficient, β , the backscatter ratio *S* is defined as $S = (\beta_{\text{aerosol}} + \beta_{\text{air}})/\beta_{\text{air}}$. Calculations are shown for prolate ice particles with an aspect ratio of 0.95. **c**, Air-parcel temperature in the wave and under synoptic conditions as from ECMWF analyses (*p* indicates pressure). **d**, Calculated surface area of liquid aerosols (thin line) and ice (thick line). The time-dependent radially resolved liquid aerosol model of Meilinger *et al.*²⁶ was used with nucleation and growth of solid particles included¹². The model was initialized with liquid particles with 9 p.p.b.v. HNO₃, 5 p.p.m.v. total H₂O and 0.49 p.p.b.v. H₂SO₄. All droplets were assumed to freeze as ice 4.1 K below the ice frost point. **e**, Chemical trajectory model results in the mountain wave (solid lines). Active chlorine is defined as $\text{Cl}_x = (2 \times \text{Cl}_2) + (2 \times \text{Cl}_2\text{O}_2) + \text{ClO}$ and is shown also for synoptic conditions (dotted line). The Mainz photochemical box model^{26,27} was used together with the particle size distributions obtained from the microphysical model. The model was initialized with 1.8 p.p.b.v. HCl, 1.2 p.p.b.v. ClONO₂, 0.015 p.p.b.v. HOCl and no Cl_x.

in the general synoptic temperature field¹². A small bias of up to 4 K in the ECMWF analysed temperatures is possible¹⁹, but this would not affect our conclusion that the ice cloud in Fig. 1a was caused by strong mesoscale cooling.

A microphysical simulation was performed along the constructed trajectory and the modelled particle sizes were used, together with a light-scattering algorithm^{12,20}, to calculate the aerosol backscatter ratio at each position through the cloud (Fig. 1b). The number of ice particles in the cloud was then adjusted to optimize agreement between the calculated and observed backscatter ratio for both polarization channels and both laser wavelengths of 532 and 354 nm. In this case, ice formation appears to have occurred in all of the liquid aerosols¹², resulting in the relatively large ice-particle surface area shown in Fig. 1d.

We have used a chemical box model to calculate the amount of chlorine activation that occurs in air parcels traversing the wave cloud (Fig. 1e). Particle surface areas in the model were taken from the results of the microphysical simulation, and all important heterogeneous reactions on liquid and ice particles were included²¹. The chemical effect of the ice cloud is clearly apparent, with ClONO₂ and HCl being rapidly converted to reactive chlorine due mainly to the reaction:



This chlorine activation is substantial, particularly considering that the ice cloud is only 15 km long and that the air spends less than 7 minutes in it. Appreciable chemical processing occurs even on the liquid aerosols, showing that ice clouds are not a prerequisite for rapid heterogeneous chlorine activation in mountain-wave events. In combination, liquid and solid chemical processing leads to production of ~1.5 parts per billion by volume, p.p.b.v., of chlorine in reactive forms (defined here as $\text{Cl}_x = (2 \times \text{Cl}_2) + (2 \times \text{Cl}_2\text{O}_2) +$

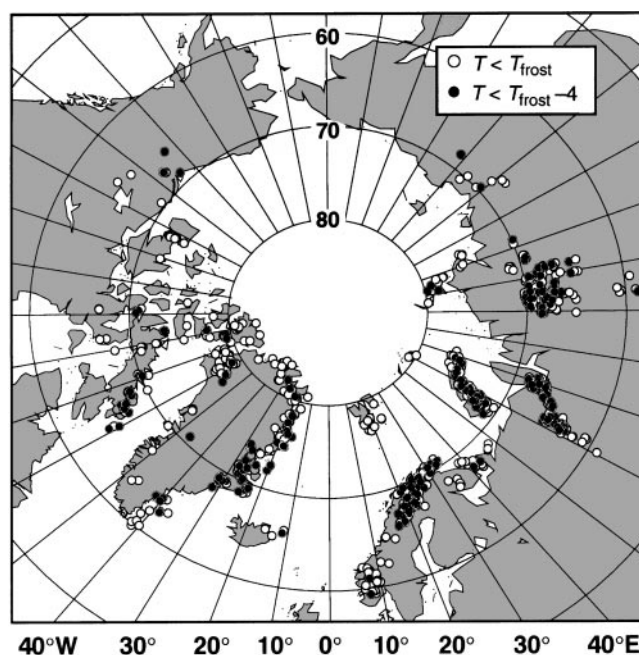
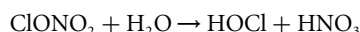


Figure 2 Predicted temperatures reached in mountain waves in the Arctic vortex during December and January 1994-95. Open circles, minimum temperatures lower than the ice frost point (T_{frost}); filled circles, minimum temperatures less than $T_{\text{frost}} - 4$ K. Calculations were made using synoptic temperatures from ECMWF analyses with mountain-wave cooling calculated using the model of Baumeister *et al.*⁸ updated daily. The calculations refer to air which was on the 550 K potential-temperature level (~25 mbar pressure) on 1 December 1994, and which sinks diabatically to 500 K by 31 January 1995.

CIO), which is almost half of the total inorganic chlorine in the air parcel. This compares with essentially no activation under equivalent synoptic conditions. It is likely that our calculations even underestimate chlorine activation in this case, as they do not include the additional processing that would occur in the larger-scale temperature minimum in which this PSC is embedded. Other PSC observations on this day show mountain-wave-induced liquid PSCs covering 200 km, and ice clouds as long as 30 km. Heterogeneous chlorine activation is complete after a single passage through these clouds. We have also examined the effect of the cloud in Fig. 1a for late-winter conditions where ClONO₂ is the principal reservoir species, with very little HCl. Chlorine activation then occurs mainly via the heterogeneous reaction



followed by photolysis of HOCl to yield Cl atoms. Heterogeneous loss of ClONO₂ was found to be comparably fast.

We have assessed the global significance of such events using a mountain-wave prediction model⁸. This model uses a global database of topographic features containing estimates of ridge height, width and orientation. A two-dimensional hydrostatic wave model driven by 5° × 2° global analyses of wind and temperature at standard pressure levels is used to calculate the adiabatic cooling above each ridge element. Maximum calculated vertical air-parcel displacements over northern Norway on the day of this PSC observation are between 1,000 and 2,000 m, equivalent to a 10–20 K change in temperature and comparable to the observed 12 K cooling derived from Fig. 1 via arguments based on microphysical

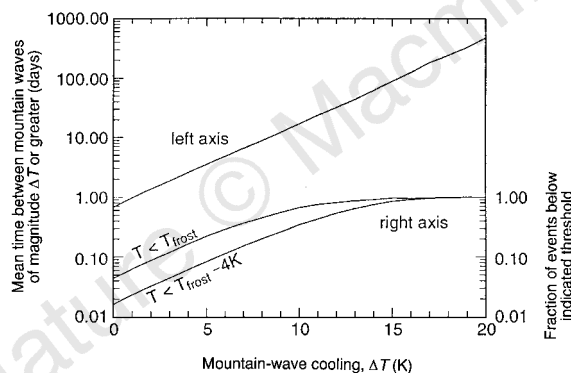


Figure 3 The frequency of occurrence of mountain-wave cooling events. Shown is the frequency with which mountain-wave cooling events greater than a given magnitude (ΔT) were encountered by air parcels confined within the Arctic polar vortex during December and January 1994–95. The results are shown as the mean time between mountain-wave events of magnitude ΔT or greater (upper curve, left axis) and the fraction of these events that lead to temperatures less than T_{frost} and $T_{\text{frost}} - 4\text{K}$ (lower curves, right axis). The simulation was performed using 190 initially evenly distributed vortex-filling trajectories initialized at 550 K potential temperature (≈ 25 mbar pressure) on 1 December 1994. Air parcels were advected using ECMWF analysed winds, sinking diabatically to ~ 500 K by 31 January 1995. All trajectories remained evenly distributed within the vortex. Temperature excursions in mountain waves were calculated according to Bacmeister *et al.*⁸ and were updated daily. The calculations refer to 16,934 wave events. The temperature excursion experienced by an air parcel is a function of the closest distance of approach to a ridge. This function is assumed here to be $\Delta T = \Delta T_{\text{max}} \exp[-(x/w)^2]$ downwind of the ridge and $\Delta T_{\text{max}} \exp[-10(x/w)^2]$ upwind, where x (in km) is the distance from the ridge line in a cross-ridge direction, w (in km) is the cross-ridge width and ΔT_{max} is the temperature change at the peak of the disturbance. This influence function represents a plausible wave-amplitude envelope close to a ridge, with most wave energy downstream. The temperature change is assumed to be zero to the left and right of the ridge.

processes. We have also examined PSC observations over Scandinavia from January 1997. These revealed ice PSCs consistent with mesoscale cooling of up to 19 K, with the MM5 model results suggesting mesoscale temperatures as much as 15 K lower than synoptic. Comparison between events such as these and the mountain-wave prediction model suggest reasonable agreement, though further efforts are needed to assess the model reliability for smaller waves.

To assess the potential effect of mountain waves for the Arctic stratosphere as a whole, we have calculated the distribution of mountain-wave cooling events within the Arctic vortex of each day of December 1994 and January 1995. We examined events that were large enough to cause significant chlorine activation—those where temperatures were forced below T_{frost} and below $T_{\text{frost}} - 4\text{K}$ (assumed to be sufficient for ice formation)—see Fig. 2. Most such events are concentrated over Norway, East Greenland and the Urals where low synoptic temperatures coincide with strong mountain-wave activity. Figure 3 shows the frequency with which mountain-wave cooling events of a given magnitude are encountered by air parcels confined within the polar vortex, as derived from diabatic synoptic trajectory calculations. Cooling events as large as shown in Fig. 1 ($\Delta T \geq 12\text{K}$) occur once every 30 days in a given air parcel, and $\sim 75\%$ of such events are calculated to lead to temperatures less than T_{frost} . Overall, temperatures in mountain waves are calculated to fall below T_{frost} approximately once every 17 days in a given air parcel, and below $T_{\text{frost}} - 4\text{K}$ once every 40 days. For the altitude level we investigated, half of the vortex air was processed below the ice frost point by 20 December, while synoptic temperatures never fell this low.

Further chemical box model calculations (not shown) reveal that the air leaving the mountain-wave PSC shown in Fig. 1 would have continued with elevated reactive chlorine concentrations for as long as 30 days, resulting in as much as 25% ozone loss in the air parcel even in the absence of further heterogeneous processing. Clearly, from the calculations shown in Fig. 3, repeated re-activation in mountain waves is likely within this time and will lead to a prolonging and intensification of the chlorine activation due to synoptic-scale PSCs. Several earlier modelling studies^{2,3} have highlighted the way in which Arctic chlorine activation and ozone destruction can be sustained by a series of relatively short-lived synoptic PSC events followed by slow reconversion of reactive chlorine species to the reservoir gases. However, in these earlier studies the time available for heterogeneous chlorine activation was of the order of a day, and thus considerably longer than the few minutes exposure to PSCs available in the mountain waves considered here. Mesoscale cooling events will be much more important in the Arctic than the Antarctic because complete chlorine activation in the Arctic due to synoptic-scale PSCs alone is uncommon, and is usually limited to a short period in mid-winter. During cold Arctic winters, such as 1994–95 studied here, the greatest effect is likely to be a prolonging of high active chlorine levels before and after the coldest periods, whereas during warmer winters the waves will induce a general increase in active chlorine concentrations through the whole winter. In both cases gravity waves will lead to enhanced ozone loss.

Several studies have indicated that chemical processing on synoptic-scale PSCs may be insufficient to explain observations of active chlorine and ozone loss in the Arctic, and a number of explanations have been proposed. Lutman *et al.*¹³ have shown that the geographical extent and peak concentrations of modelled ClO in the Arctic during 1991–92 were persistently lower than measured by the satellite-borne Microwave Limb Sounder (MLS). Bell *et al.*¹⁴ also report a model underprediction of ClO for the winter 1994–95, while Douglass *et al.*¹⁵ found that the early onset of elevated ClO levels observed by MLS during December 1991 could be reproduced in their model only by reducing synoptic temperatures by 2 K overall.

Consistent with low modelled reactive chlorine concentrations, several workers report model underpredictions of ozone depletion of between 25 and 40% (refs 16, 17), although these discrepancies may be due partly to incorrectly modelled gas-phase reactions rather than heterogeneous processing. Edouard *et al.*²² have suggested that such underprediction by three-dimensional models may be caused by averaging small-scale structure in species concentration over large grid boxes. This arises because of the nonlinear dependence of key reaction rates on concentration. This offers a possible explanation for underpredicted ozone loss in three-dimensional models, but does not explain low modelled ClO and does not refer to predictions of trajectory models¹³. Whether uncertainties in the temperature fields used in models are responsible for model underpredictions^{13,15}, or whether the mountain-wave activation mechanism suggested here is inappropriate can only be assessed by undertaking a multi-year model simulation and comparison with available data. However, it should be noted that mountain waves can act as a continuous source of inhomogeneities in reactive chlorine concentrations on a scale of a few kilometres, which, according to the findings of Edouard *et al.*²², will make their inclusion in three-dimensional models highly problematic.

Other sources of gravity waves should also be examined for their potential role in chlorine activation; Wu and Waters^{23,24} have shown from satellite data that there is evidence for middle-atmosphere gravity waves that correlate not only with surface topography but also with upper-tropospheric convection and stratospheric polar vortices. Gravity waves occurring at the vortex edge should also be examined for their potential contribution to the general decrease in ozone which has been observed at middle latitudes over the past decade¹. □

Received 30 July 1997; accepted 8 January 1998.

1. Scientific Assessment of Ozone Depletion: 1994 (Rep. No. 37, World Meteorological Organisation, Geneva, 1994).
2. Jones, R. L., McKenna, D. S., Poole, L. R. & Solomon, S. On the influence of polar stratospheric cloud formation on chemical composition during the 1988/89 Arctic winter. *Geophys. Res. Lett.* **17**, 545–548 (1990).
3. McKenna, D. S. *et al.* Calculations of ozone destruction during the 1988/89 Arctic winter. *Geophys. Res. Lett.* **17**, 553–556 (1990).
4. Borrmann, S. *et al.* Heterogeneous reactions on stratospheric background aerosols, volcanic sulfuric acid droplets, and type 1 polar stratospheric clouds. *J. Geophys. Res.* **102**, 3639–3648 (1997).
5. Gary, B. L. Observational results using the microwave temperature profiler during the Airborne Antarctic Ozone Experiment. *J. Geophys. Res.* **94**, 11223–11231 (1989).
6. Volkert, H. & Intes, D. Orographically forced stratospheric waves over northern Scandinavia. *Geophys. Res. Lett.* **19**, 1205–1208 (1992).
7. Leutbecher, M. & Volkert, H. Stratospheric temperature anomalies and mountain waves: A three dimensional simulation using a multi-scale weather prediction model. *Geophys. Res. Lett.* **23**, 3329–3332 (1996).
8. Bacmeister, J. T., Newman, P. A., Gary, B. L. & Chan, K. R. An algorithm for forecasting mountain wave-related turbulence in the stratosphere. *Weather Forecast.* **9**, 241–253 (1994).
9. Dörnbrack, A., Leutbecher, M., Volkert, H. & Wirth, M. Mesoscale forecasts of stratospheric mountain waves. *Meteorol. Appl.* (in the press).
10. Stanford, J. L. & Davis, J. S. A century of stratospheric cloud reports: 1870–1972. *Bull. Am. Meteorol. Soc.* **55**, 213–219 (1974).
11. Peter, Th., Müller, R., Rrdla, K., Petzoldt, K. & Reimer, E. A microphysical box model for EASOE: Preliminary results for the January/February 1990 PSC event over Kiruna. *Ber. Bunsenges. Phys. Chem.* **96**, 362–367 (1992).
12. Carslaw, K. S. *et al.* Particle microphysics and chemistry in remotely observed mountain polar stratospheric clouds. *J. Geophys. Res.* (in the press).
13. Lutman, E. R. *et al.* Three-dimensional studies of the 1991/92 northern hemisphere winter using domain-filling trajectories with chemistry. *J. Geophys. Res.* **102**, 1479–1488 (1997).
14. Bell, W. *et al.* Measurements of stratospheric chlorine monoxide (ClO) from ground-based FTIR observations. *J. Atmos. Chem.* **24**, 285–297 (1996).
15. Douglass, A. *et al.* A three-dimensional simulation of the early winter distribution of reactive chlorine in the north polar vortex. *Geophys. Res. Lett.* **20**, 1271–1274 (1993).
16. Goutail, F. *et al.* in *Polar Stratospheric Ozone* (eds Pyle, J. A., Harris, N. P. & Amanatidis, G. T.) 574–579 (Air Pollution Res. Rep. No. 56, European Communities, Brussels, 1996).
17. Hansen, G., Svenoe, T., Chipperfield, M. P., Dahlback, A. & Hoppe, U.-P. Evidence of substantial ozone depletion in winter 1995/96 over northern Norway. *Geophys. Res. Lett.* **24**, 799–802 (1997).
18. Carslaw, K. S., Peter, Th. & Clegg, S. L. Modeling the composition of liquid stratospheric aerosols. *Rev. Geophys.* **35**, 125–154 (1997).
19. Knudsen, B. M. Accuracy of Arctic stratospheric temperature analyses and the implications for the prediction of polar stratospheric clouds. *Geophys. Res. Lett.* **23**, 3747–3750 (1996).
20. Mishchenko, M. I. Light scattering by randomly oriented axially symmetric particles. *J. Opt. Soc. Am.* **8**, 871–882 (1991).
21. DeMore, W. *et al.* *Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling* (Publ. 97-4, Jet Propulsion Lab., Pasadena, 1997).
22. Edouard, S., Legras, B., Lefevre, F. & Eymard, R. The effect of small-scale inhomogeneities on ozone depletion in the Arctic. *Nature* **384**, 444–447 (1996).
23. Wu, D. L. & Waters, J. W. Gravity-wave-scale temperature fluctuations seen by the UARS MLS. *Geophys. Res. Lett.* **23**, 3289–3292 (1996).

24. Wu, D. L. & Waters, J. W. Satellite observations of atmospheric variances—a possible indication of gravity waves. *Geophys. Res. Lett.* **24**, 3631–3634 (1996).
25. Meilinger, S. K. *et al.* Size-dependent stratospheric droplet composition in lee wave temperature fluctuations and their potential role in PSC freezing. *Geophys. Res. Lett.* **22**, 3031–3034 (1995).
26. Müller, R. *et al.* Chlorine chemistry and the potential for ozone depletion in the Arctic stratosphere in the winter 1991/92. *Geophys. Res. Lett.* **21**, 1427–1430 (1994).
27. Carslaw, K. S., Peter, Th. & Müller, R. Uncertainties in reactive uptake coefficients for solid stratospheric particles—2. Effect on ozone depletion. *Geophys. Res. Lett.* **24**, 1747–1750 (1997).

Acknowledgements. We thank M. Mishchenko for help with the T-matrix calculations. Part of this work was funded by the European Communities, the German BMBF, the Office of Naval Research (ONR) and NASA's Atmospheric Chemistry Modeling and Analysis Program (ACMAP).

Correspondence and requests for materials should be sent to K.S.C. (e-mail: carslaw@mpch-mainz.mpg.de).

Reduced sensitivity of recent tree-growth to temperature at high northern latitudes

K. R. Briffa*, F. H. Schweingruber†, P. D. Jones*, T. J. Osborn*, S. G. Shiyatov‡ & E. A. Vaganov§

* Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK

† Swiss Federal Institute of Forest, Snow and Landscape Research, Zürcherstrasse 111, CH-8903, Birmensdorf, Switzerland

‡ Institute of Plant and Animal Ecology, Ural Branch of the Russian Academy of Sciences, 8 Marta Street, Ekaterinburg 620219, Russia

§ Institute of Forest, Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russia

Tree-ring chronologies that represent annual changes in the density of wood formed during the late summer can provide a proxy for local summertime air temperature¹. Here we undertake an examination of large-regional-scale wood-density/air-temperature relationships using measurements from hundreds of sites at high latitudes in the Northern Hemisphere. When averaged over large areas of northern America and Eurasia, tree-ring density series display a strong coherence with summer temperature measurements averaged over the same areas, demonstrating the ability of this proxy to portray mean temperature changes over sub-continents and even the whole Northern Hemisphere. During the second half of the twentieth century, the decadal-scale trends in wood density and summer temperatures have increasingly diverged as wood density has progressively fallen. The cause of this increasing insensitivity of wood density to temperature changes is not known, but if it is not taken into account in dendroclimatic reconstructions, past temperatures could be overestimated. Moreover, the recent reduction in the response of trees to air-temperature changes would mean that estimates of future atmospheric CO₂ concentrations, based on carbon-cycle models that are uniformly sensitive to high-latitude warming, could be too low.

We have based this analysis on tree-growth data derived from trees sampled at more than 300 locations spread across the Northern Hemisphere (Fig. 1). These generally cool and moist sites were all chosen at relatively high latitudes or high elevations for the potential sensitivity of year-by-year tree growth to local temperature variability². Though many of the sites are not strictly at tree-line, these data are a good proxy for the growth of the northern part of the boreal forest, rather than the large areas of closed-canopy, more drought-sensitive southern coniferous forests. Ring widths and the densities of the parts of the rings formed in late summer (maximum-latewood ring densities) were measured using multiple tree cores from each site¹. The measurement series were then individually detrended to remove inherent age-related bias, by taking residuals from a simple function fit to the raw data and the resulting dimensionless indices averaged to form separate ring-width and

density chronologies for each site³. The detrending process results in a loss of multi-century variance, the extent of which is dependent on tree age and the flexibility of the fitted function^{4,5}. These chronologies represent tree growth variability on timescales of years to centuries.

The fidelity of the local temperature sensitivity of these data has been clearly demonstrated by comparison with instrumental records and by the use of similar chronologies, or groups of chronologies, to reconstruct either regional-mean or detailed spatial patterns of past temperature variability at various locations, such as in northern Fennoscandia, western and northern North America and northern Russia^{2,6–9}, frequently using principal component regression techniques¹⁰. Here we have simply combined the dimen-

sionless chronologies into large regional mean series, as defined in Fig. 1 legend, and rescaled the regional means and variances to match over a common base (1881–1940). Our purpose is to identify very large (sub-continental scale) growth signals, necessarily obscuring any individual and localized site factors.

Table 1a shows the strength of the associations between these regional chronologies and summer mean temperatures¹¹ (April–September for density and June–August for the ring-width comparisons) averaged over similar areas (Fig. 1). The correlations focus separately on short (interannual) and longer-timescale (decadal and above) common variability in that they are calculated using high-pass (<10-year) and low-pass (>10-year) filtered data respectively.

On the interannual timescale, averaged across all regions, about

Table 1 Correlations between tree-ring and temperature series

	Maximum-latewood density*				Ring widths†			
	Interannual‡		Decadal§		Interannual		Decadal	
	1881–1960	1881–1981	1881–1960	1881–1981	1881–1960	1881–1981	1881–1960	1881–1981
a Regional-mean tree-growth series versus temperature								
NWNA	71	66	58	46	35	33	12	15
SWNA	72	75	72	63	27	22	–4	4
ENA	72	73	82	70	33	31	26	15
NEUR	89	87	92	81	73	64	56	45
SEUR	59	62	82	79	44	39	–6	7
WSIB	82	83	75	58	41	42	45	28
CSIB	56	51	83	67	42	34	54	44
ESIB	52	56	78	52	27	37	21	37
NORTH	73	71	90	60	32	28	78	34
SOUTH	64	67	73	65	37	41	12	13
ALL	68	70	89	64	47	47	74	43
b Tree-growth series versus Northern Hemisphere mean temperatures								
NORTH	45	44	74	41	16	11	79	42
SOUTH	2	10	61	40	8	14	25	28
ALL	33	36	81	46	19	17	75	51

All correlations in this table have been multiplied by 100.

* Correlated against April–September mean instrumental averages.

† Correlated against June–August instrumental data.

‡ Interannual correlations involve timeseries of residuals from decadal smoothed curves.

§ Decadal correlations are calculated using decadal smoothed series.

|| Regional acronyms defined in Fig. 1 legend.

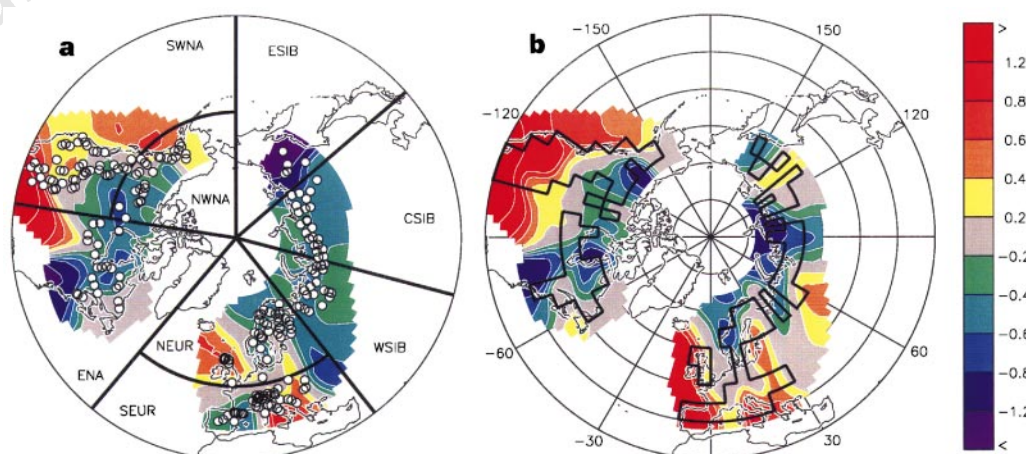


Figure 1 Spatial patterns of relative tree-growth decline. **a**, The location (circles) of tree-ring chronologies and the division (black lines) into regional averages. **b**, The locations (black lines) of the grid-box temperatures used for comparison with the tree growth series. The coloured contours show where the ring density (**a**) and ring width (**b**) are enhanced (positive) or suppressed (negative) relative to summer temperature during the period 1975–85 compared to the period 1935–45. Values are dimensionless, being the change in the difference of two normalized

time series in each grid box. Definition of regions and number of sites: southwestern (SWNA, 53 sites), northwestern (NWNA, 30) and eastern (ENA, 34) North America; northern (NEUR, 46) and southern (SEUR, 72) Europe; western (WSIB, 42), central (CSIB, 31) and eastern (ESIB, 6) Siberia; all 125 sites in SWNA and SEUR form the composite region SOUTH, and all 189 sites in the six northern regions form the NORTH region; ALL is an average of all 314 sites.

50% of the variance (r^2) of each temperature series is common to the density series, with the individual regional variances ranging from 80% (NEUR) to ~30% (CSIB and ESIB). Consistently similar correlations are achieved, either calculated over the shorter (1881–1960) or extended (to 1981) periods, or when larger regional density chronologies made up of the NORTH (incorporating the NRNA, ENA, NEUR, WSI, CSIB and ESIB) and SOUTH (SWNA and SEUR) data are compared with the equivalent regionally-averaged temperatures.

The regional correlations for the decadal smoothed densities and temperatures are, in general, higher than the interannual values, with per cent common variances ranging from 34 (NRNA) to 85 (NEUR) and averaging ~60%: but only for the period 1881–1960. When the decadal correlations are calculated over the longer 1881–1981 period, consistent falls in the common variances are apparent in all areas. The falls are relatively small in the southern regions (from 67% to 62% in SEUR; and from 52% to 40% in SWNA) but much greater in the northern areas with a maximum fall in ESIB of >30% (from 61% to 27%). The average degradation in decadal-timescale common variance is nearly 20% across all eight regions. This low-frequency loss in temperature sensitivity is even more apparent when the correlations for the larger regions are compared over the two periods: over 30% loss of common variance overall, with the simple correlations falling from 0.90 to 0.60 in the NORTH and from 0.73 to 0.65 in the SOUTH, and 0.89 to 0.64 in ALL (the average series for all 314 sites).

For the period before 1960, these smoothed NORTH and SOUTH density averages also have 55% and 37% variance in common, respectively, with the full 'Northern Hemisphere' (land and marine) mean temperature record¹¹ (NHT), while density series ALL has 66% variance in common with the NHT (Table 1b). (The

latter can be compared with a value of 87% for the comparison between the actual temperature data for the 'ALL' region and the NHT series). Recalculating these density and NHT correlations over a period extended by only 21 years results in common variances falling by 17% and 16% for the NORTH and SOUTH series, and 21% for the ALL.

In general, the ring-width/temperature correlations are lower than those for the densities. This is particularly evident on an interannual timescale because ring widths are more strongly auto-correlated and may integrate temperature forcing over a wider, seasonal (perhaps annual) window and over more than a single year^{9,12,13}. However, the recent relative loss of decadal temperature sensitivity is also apparent in the ring-width data and is more noticeable in the larger spatial aggregate series where the temperature associations are stronger, especially, as with density, in the northern regions.

Figure 2 illustrates the temporal changes in the association between the smoothed tree growth and appropriate temperature series for the eight individual regions and for the NORTH, SOUTH and ALL aggregate series. Instrumental summer temperatures show slight increases in all areas. Most of this warming occurred in the first half of this century so that mean temperature levels in recent decades and the warmth of the late 1980s only match, or barely exceed, the levels of earlier warm periods. A negative bias in the comparable tree-growth records during the second half of this century is apparent in all regions. In all but the two southern regions, the temperature and tree growth curves increasingly diverge. In the areas where the growth data extend through to the warm late 1980s and early 1990s (NEUR, WSIB, CSIB, ESIB), the divergence is at a maximum in the most recent years.

Averaged around the Northern Hemisphere, early tree growth

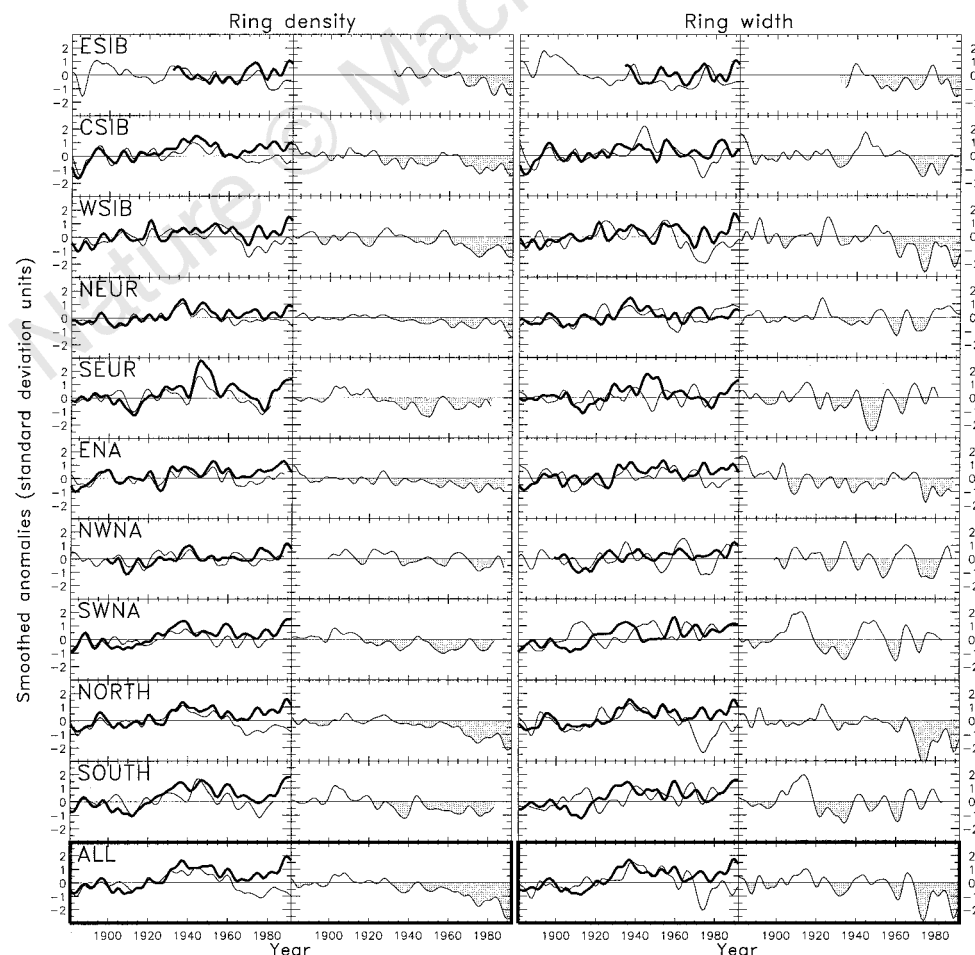


Figure 2 Regional tree growth and temperatures over the past 120 years. Decadal smoothed tree growth (thin lines), maximum-latewood density or ring width, plotted against mean summer temperatures (thick lines), April–September for density and June–August for ring width, for each of the regions described in Fig. 1. The difference series (growth minus temperature), shaded to emphasize negative values, are shown to the right of each pair of curves. All data series have been scaled to have zero mean and unit variance over the period 1881–1940 (except the short ESIB temperature series which uses 1932–75).

(ALL) can be seen to follow closely the decadal trends in recorded summer temperatures, tracking the rise to the relatively high levels of the 1930s and 1940s and the subsequent fall in the 1950s. However, although temperatures rose again after the mid-1960s and reached unprecedentedly high recorded levels by the late 1980s, hemispheric tree growth fell consistently after 1940, and in the late 1970s and 1980s reached levels as low as those attained in the cool 1880s. Over the hemisphere, the divergence between tree growth and mean summer temperatures began perhaps as early as the 1930s; became clearly recognisable, particularly in the north, after 1960; and has continued to increase up until the end of the common record at around 1990. The reason for this increasingly apparent and widespread phenomenon is not known but any one, or a combination, of several factors might be involved.

The positive linear thermal response of the trees may break down above some absolute threshold, perhaps with lower soil moisture exerting increasing stress at a number of sites¹³. However, the fidelity of the long-term growth response to the relatively high temperatures in the middle part of the century would imply the need for a concomitant lowering of soil moisture levels over very widespread areas during recent decades. There is little evidence that this has occurred in the regions represented by these trees^{14,15}. Also the maintenance of high positive interannual temperature correlations during recent decades, argues against a simple soil-moisture-related explanation of the reducing tree growth. However, more subtle changes in precipitation or temperature seasonality, such as the documented move to warmer springs, and associated earlier snow melt¹⁶ mean that it would be unwise to dismiss summer drought sensitivity too readily. Other factors such as increasing competition with other plants and increasing insect herbivory could play a role^{17,18}, but the apparent widespread synchronicity of the phenomenon suggests a hemispheric-scale influence. Higher UV-B levels^{19–21} or decreased solar radiation receipts (increased optical depth)^{22–24} may be involved, but attempting to isolate such effects will be problematic in the presence of increasing atmospheric CO₂ concentrations and increasing amounts of acidic deposition and tropospheric ozone, all of which, along with climate variability, are likely to influence tree growth^{25,26}.

Whatever the cause, this change in tree-growth response has important implications for studies of past and future climate change. The common use of least-squares regression for developing dendroclimatic transfer function equations to estimate past climates, imposes an equality of means in both the predictand and predictor time series over the fitting or 'calibration' period¹⁰. Any bias in mean tree growth will, therefore, be 'corrected' during calibration, with the consequence that the derived regression coefficients will be biased. Our results imply that this might increasingly result in systematic overestimation of past temperatures, particularly in regions where the loss of low-frequency temperature sensitivity in tree growth is greatest (eastern Siberia and eastern North America: see Fig. 1), where tree-ring 'standardization' is designed to preserve maximum long-timescale chronology variability⁴, and where transfer functions are calibrated over the most recent decades.

The potential importance of these results for modelling future temperatures derives from their relevance to carbon-cycle-model experiments that seek to explain the history of measured atmospheric CO₂ concentrations and provide realistic projections^{27,28}. Along with spatio-temporal analyses of past CO₂ measurements, these experiments show the existence of an important missing CO₂ sink, a large part of which is believed to be the northern terrestrial biosphere, including the boreal forest²⁹. Accurate parametrization of the magnitude of this sink is necessary in order to reduce the uncertainties in the CO₂ mass balance and hence in projections of future CO₂ concentrations and climate²⁷. Although we emphasize that our results apply generally only to the high-latitude temperature-sensitive forest regions of the Northern Hemisphere, the degrada-

tion in thermal response of this significant area of the world's biosphere suggests that we should be cautious in assuming, in carbon-cycle-model experiments, a constant temperature-dependent biospheric CO₂ uptake during different halves of the present century, and in any future warmer world. Such an assumption may lead to underestimates of future atmospheric concentrations of CO₂ and conservative estimates of future warming. However, the real significance of our results, in this regard, can only be gauged by quantifying the impact of reduced tree growth on absolute amounts of carbon sequestered and comparing these with estimates over previous centuries based on earlier tree-ring data. □

Methods

The individual site tree-ring chronologies were produced from multiple increment core samples replicated within and between trees. The individual ring width (TRW) and maximum-latewood-density (MXD) measurement series were corrected for age-related bias by taking residuals from a generalized exponential or "Hugershoff" function of the form $G_t = at^b \exp(-ct)$, where G_t is the biological growth trend, t is time and a , b and c are empirically fitted constants³. A simple arithmetic average of the resulting dimensionless indices is computed across each year to form each site chronology. We examined a cross-section of the data looking for systematic misfits of the function in recent years without finding evidence that the results we describe are an artefact of inappropriate standardization.

Before measuring absolute densities, resin and heartwood compounds were extracted with alcohol and water. There was no evidence of heartwood/sapwood bias in the density profiles and no evidence of bias associated with technical difficulties measuring very narrow rings.

The individual MXD and TRW site chronologies were normalized with respect to their mean and standard deviation for the period 1901–40. Within each region, all chronologies were averaged together with a time-dependent weighting equal to the fraction of all possible tree cores that were available for a particular year of each chronology. Each regional mean thus obtained tended to have greater variance during years when few chronologies were available to contribute to the average; this effect was corrected³⁰ for by scaling by the square root of the effective number (n') of independent samples available in each year, where $n' = n/[1 + (n-1)\bar{r}]$. Here, $\bar{r}$ is the mean intersite correlation between the n chronologies, a measure of the regional-scale common signal. After correction, each regional mean timeseries was normalized with respect to 1881–1940.

Regional mean temperature time series were constructed by averaging together the stated monthly mean temperature anomalies from each 5° by 5° box in which a chronology for that region is located (Fig. 1). The variance of each of these series was also corrected according to the effective number of independent samples that contributed to each year's mean, and was also normalized with respect to 1881–1940. The Northern Hemisphere mean temperature series is made up of all available gridbox data, including land and marine data, north of the Equator.

In the most recent period (1983–92), the regional mean tree-growth timeseries (particularly ALL and NORTH) are spatially biased towards northern Eurasia, due to the earlier development of chronologies in other regions. For comparison, similarly biased temperature time series were constructed by using a time-dependent sampling area defined each year by choosing only those boxes that had MXD and TRW chronology values in them. These biased series were very similar to those temperatures computed from the fixed regional grids (Fig. 1) and do not alter any of the conclusions drawn here.

Received 14 May; accepted 11 November 1997.

1. Schweingruber, F. H. *Tree Rings. Basics and Applications of Dendrochronology* (Kluwer, Dordrecht, 1988).
2. Schweingruber, F. H. & Briffa, K. R. in *Climate Variations and Forcing Mechanisms of the Last 2000 Years* (eds Jones, P. D., Bradley, R. S. & Jouzel, J.) 43–66 (Springer, Berlin, 1996).
3. Bräker, O. Der Alterstrend bei Jahrringdichten und Jahrringbreiten von Nadelhölzern und sein Ausgleich. *Mitt. forstl. Bundesvers. Anst. Wien* **142**, 75–102 (1981).
4. Briffa, K. R., Jones, P. D., Schweingruber, F. H., Karlén, W. & Shiyatov, S. G. in *Climate Variations and Forcing Mechanisms of the Last 2000 Years* (eds Jones, P. D., Bradley, R. S. & Jouzel, J.) 9–41 (Springer, Berlin, 1996).
5. Cook, E. R., Briffa, K. R., Meko, D. M., Graybill, D. A. & Funkhouser, G. The 'segment-length curse' in long tree-ring chronology development for palaeoclimatic studies. *Holocene* **5**, 229–237 (1995).
6. Schweingruber, F. H., Briffa, K. R. & Nogler, P. A tree-ring densitometric transect from Alaska to Labrador. *Int. J. Biometeorol.* **37**, 151–169 (1993).

7. D'Arrigo, R. D., Jacoby, G. C. & Free, R. M. Tree-ring width and maximum latewood density at the North American tree line: parameters of climate change. *Can. J. Forest Res.* **22**, 1290–1296 (1992).
8. Graybill, D. A. & Shiyatov, S. G. in *Climate Since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 393–414 (Routledge, London, 1992).
9. Graumlich, L. J. A 1000-year record of temperature and precipitation in the Sierra Nevada. *Quat. Res.* **39**, 249–255 (1993).
10. Cook, E. R., Briffa, K. R. & Jones, P. D. Spatial regression methods in dendroclimatology: a review and comparison of two techniques. *Int. J. Climatol.* **14**, 379–402 (1994).
11. Jones, P. D. & Briffa, K. R. Global surface air temperature variations during the twentieth century: Part 1, spatial, temporal and seasonal details. *Holocene* **2**, 165–179 (1992).
12. Jacoby, G. C. & D'Arrigo, R. D. Reconstructed Northern Hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America. *Clim. Change* **14**, 39–49 (1989).
13. Jacoby, G. C. & D'Arrigo, R. D. Tree ring width and density evidence of climatic and potential forest change in Alaska. *Glob. Biogeochem. Cycles* **9**, 227–234 (1995).
14. Briffa, K. R., Jones, P. D. & Hulme, M. Summer moisture variability across Europe, 1892–1991: an analysis based on the Palmer Drought Severity Index. *Int. J. Climatol.* **14**, 475–506 (1994).
15. Nicholls, N. *et al.* in *Climate Change 1995* (eds Houghton, J. T. *et al.*) 133–192 (Cambridge Univ. Press, 1996).
16. Groisman, P. Y., Karl, T. R. & Knight, R. W. Observed impact of snow cover on the heat balance and the rise of continental spring temperatures. *Science* **263**, 198–200 (1994).
17. Hogg, P. H. & Hurdle, P. A. The aspen parkland in western Canada: a dry-climate analogue for the future boreal forest? *Wat. Air Soil Pollut.* **82**, 391–400 (1995).
18. Fleming, R. A. & Volney, J. A. Effects of climate change on insect defoliator population processes in Canada's boreal forest: some plausible scenarios. *Wat. Air Soil Pollut.* **82**, 445–454 (1995).
19. DeLuise, J. J. *et al.* Northern and middle-latitude ozone profile features and trends observed by SBUV and Umkehr, 1979–1990. *J. Geophys. Res.* **99**, 18901–18908 (1994).
20. Sullivan, J. H. in *Stratospheric Ozone Depletion/UV-B Radiation in the Biosphere* (eds Biggs, R. H. & Joyner, M. E. B.) 67–76 (Springer, Berlin, 1994).
21. Tevini, M. in *Stratospheric Ozone Depletion/UV-B Radiation in the Biosphere* (eds Biggs, R. H. & Joyner, M. E. B.) 37–56 (Springer, Berlin, 1994).
22. Bryson, R. A. & Goodman, B. M. Volcanic activity and climatic changes. *Science* **207**, 1041–1044 (1980).
23. Asaturov, M. L. *et al.* *Volcanoes, Stratospheric Aerosols and Earth's Climate* (Gidrometeoizdat, Leningrad, 1986) [In Russian].
24. Bradley, R. S. & Jones, P. D. in *Climate Since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 606–622 (Routledge, London, 1992).
25. Melillo, J. M., Prentice, I. C., Farquhar, G. D., Schulz, E.-D. & Sala, O. E. in *Climate Change 1995* (eds Houghton, J. T. *et al.*) 444–481 (Cambridge Univ. Press, 1996).
26. Hättenschwiler, S., Schweingruber, F. H. & Körner, C. Tree ring responses to elevated CO₂ and increased N deposition in *Picea Abies*. *Plant Cell Environ.* **19**, 1369–1378 (1996).
27. Wigley, T. M. L. Balancing the carbon budget. Implications for projections of future carbon dioxide concentration changes. *Tellus B* **45**, 409–425 (1993).
28. Craig, S. G. & Holmén, K. J. Uncertainties in future CO₂ projections. *Glob. Biogeochem. Cycles* **9**, 139–152 (1995).
29. Ciais, P., Tans, P. P., Troler, M., White, J. W. C. & Francey, R. J. A large northern hemisphere terrestrial CO₂ sink indicated by the ¹³C/¹²C ratio of atmospheric CO₂. *Science* **269**, 1098–1102 (1995).
30. Osborn, T. J., Briffa, K. R. & Jones, P. D. Adjusting variance for sample size in tree-ring chronologies and other regional-mean timeseries. *Dendrochronologia* (in the press).

Acknowledgements. We thank P. Nogler, T. Forster, B. Feiertag and E. Schär, WSLFNP, for densitometry and chronology construction, and G. Jacoby and C. Körner for comments on the manuscript. This work was supported by the Swiss Federal Institute of Forest, Snow and Landscape Research, the Swiss National Science Foundation, the European Community Environment and Climate Programme and the UK NERC.

Correspondence and requests for materials should be addressed to K.R.B. (e-mail: k.briffa@uea.ac.uk).

Evidence for partial melt at the core–mantle boundary north of Tonga from the strong scattering of seismic waves

John E. Vidale* & Michael A. H. Hedlin†

* Department of Earth and Space Sciences, University of California, Los Angeles, Los Angeles, California 90095-1567, USA

† Cecil H. and Ida M. Green Institute of Geophysics and Planetary Physics, University of California, San Diego, La Jolla, California 92093-0225, USA

Scattered waves that precede the seismic phase PKP (which traverses the Earth's core) have been used to identify and locate small-scale heterogeneity in the Earth's mantle^{1–6}. A recent study has demonstrated that the global data set of these precursors is consistent with weak heterogeneity (about 1 per cent r.m.s. velocity variation) distributed throughout the mantle⁷. Here we show, however, that anomalously large PKP precursors from earthquakes in northern Tonga require much stronger heterogeneity (10–15 per cent r.m.s. velocity variation) in a layer about

60 km thick near the core–mantle boundary below Tonga. This region of the core–mantle boundary is also marked by low shear-wave velocities in the lower mantle⁸ and is near an area of very low compressional-wave velocity in the lowermost tens of kilometres of the mantle⁹, which has been interpreted as evidence for the presence of partial melt¹⁰. The strength of the scattering that we observe provides strong support for the presence of partial melt in this area, and also suggests that vigorous small-scale convection is taking place at the core–mantle boundary.

The nature of the lowermost mantle remains elusive. As the core–mantle boundary (CMB) is the lower bound on mantle convection, it must underlie at least a thermal boundary layer¹¹. However, just above the CMB may also be home to chemical heterogeneity¹². The lowermost mantle shows two kinds of seismic structure. First, layering in P- and S-wave velocities and anisotropy is seen in the few hundred kilometres above the CMB¹³. Second, weak fine-scale heterogeneity is inferred from scattered waves that precede the seismic phase PKP^{1–6}. Until recently, the few per cent amplitudes inferred for these two types of structures were small enough that variations in mineralogy and reasonable temperature differences without melt could provide an explanation^{7,11,14}. One or more patches of the CMB, which correlate with large-scale slower than average lower mantle, seem to have greater perturbations. Recent work^{9,15} suggests P-wave velocity reductions of ~10% in the lowest

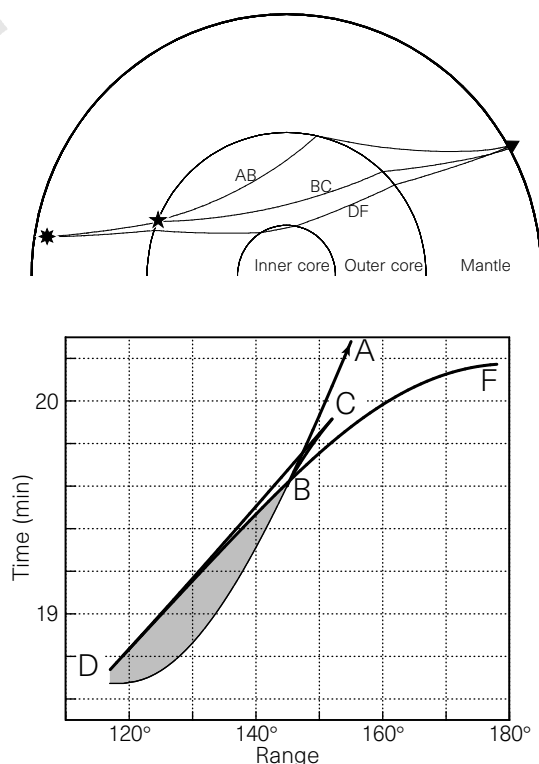


Figure 1 Ray-paths and travel times of PKP phases. A receiver located 140° from an earthquake can detect seismic energy that has passed through the inner core (PKP_{at}) as well as energy deflected from P to the outer core branches PKP_{ab} and PKP_{bc} by a scatterer located at the CMB. Owing to the unusual ray geometries caused by the sharp drop in P-wave velocity at the CMB, this scattered energy can precede PKP_{at}. Although we have depicted near-source scattering at the CMB, scattering at any depth in the mantle at either end of the path can give rise to precursors. The lower panel shows the four branches of PKP from a surface event. In addition to the inner and outer core refracted phases just discussed is PKP_{cd}, the phase that reflects off the inner core. Possible arrival times for the precursors are indicated by the shaded region. In the top panel, the eight-pointed star marks the earthquake, the five-pointed star marks a possible scattering region, and the triangle marks the receiver.

tens of kilometres of the mantle. Such dramatic reductions may indicate the presence of partial melt in the mantle next to the core¹⁰. Here we present observations of scattered PKP waves requiring large velocity perturbations near the CMB, and note their position beneath low-velocity lower mantle and near a patch exhibiting very slow basal P velocities.

The path that PKP precursors travel through the Earth, by scattering of P waves in the mantle, has been recognized for many years (refs 1–6; Fig. 1). Global surveys of PKP precursors reveal generally weak scattering. For example, near a distance of 136° from earthquakes, within the distance range we analyse, the scattered precursor phases tend to range from 0 to 20% of the amplitude of PKP on short-period recordings^{7,16}. The data we present below, however, show much stronger precursors.

In Fig. 2 we show the locations of 22 earthquakes from Tonga and New Hebrides that were recorded on the NORSAR array in Norway. All events were deeper than 100 km and had a moment greater than 10^{26} dyne-cm. The duration of faulting for all events has been measured to range from 4 to 10 s in a study of the rupture properties of deep earthquakes¹⁷. Three eligible events were discarded; one recording was too noisy, and the other two earthquakes had gradual beginnings that were comparable to the interval between PKP and its precursors.

Anomalous PKP precursors are shown in Fig. 3a and b. The precursors are nearly as large as the PKP waves, and lack long-period energy. The precursors and PKP waves arrive with differing apparent velocities, and the precursors are not as coherent as the PKP waves. The time function of rupture in the earthquake, which may be seen in Fig. 3c, does not explain the complexity of the observed stack. Stacks from the 22 events are shown in Fig. 4. It is clear that the precursor becomes so large beyond 138° that the PKP phase becomes difficult to identify. Precursors as close as 136° have amplitudes comparable to PKP.

PKP precursors in a single seismogram may arise from scattering in the mantle either below the source or the receiver¹⁶; for example, an event from Tonga recorded at NORSAR (event 7 in ref. 6) showed receiver-side precursors. In our case, the region of the mantle that contributes to source-side scattering for Tonga and New Hebrides has some overlap, while the region contributing to receiver-side scattering is nearly identical. Earthquakes in the New Hebrides subduction zone have smaller scattered PKP precursors than earthquakes from Tonga. The largest precursors clearly have greater

slowness than PKP. These observations indicate predominantly source-side scattering for the Tonga earthquakes, as had been previously inferred from similar data¹⁸.

There is a remote possibility that structure in the inner core is removing the high-frequency component of PKP (see Fig. 1 for nomenclature) and misleading us into overestimating the strength of the precursors. If so, the inner-core structure would have to be highly anomalous, as comparable precursors are not observed for other paths.

The anomalous scattering zone seen in the records of the Tonga events is geographically limited. PKP precursors sampling neighbouring areas do not show nearly such large precursor amplitudes. The recordings of the events in New Hebrides show smaller precursors than Tongan events, although still bigger precursors than the global average. French nuclear tests in the Tuamotu archipelago 3,000 km to the east show PKP precursors with normal amplitudes. In addition, the strongest precursors arrive as a distinct pulse in time, whereas scattering from a broader zone would produce a more prolonged precursor wavetrain that would not have a minimum between the precursors and PKP.

A possible complication is that the presence of a zone of strong scattering near the CMB could attenuate the PKP phase as well as amplify the precursors in our seismograms. However, we doubt that PKP is greatly attenuated for the following reasons. First, the amplitudes of the PKP phases from Tonga are similar to those of the PKP arrivals from New Hebrides. Second, attenuation of PKP should vary more with earthquake location than generation of precursors, which is not observed.

We model these arrivals as caused by scattering from a random

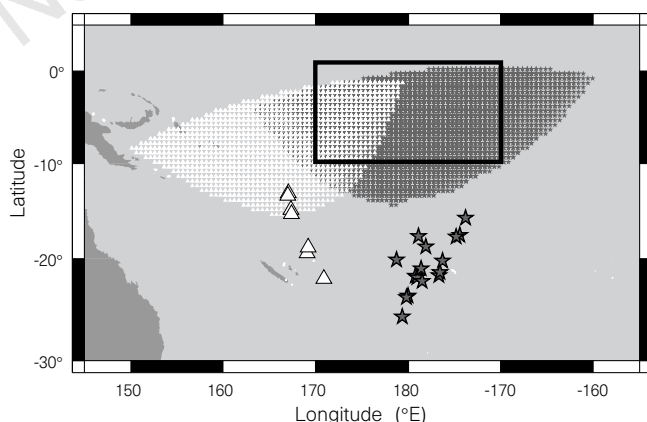


Figure 2 Map of anomalous region of the CMB and the 22 earthquakes used in this study. Dark grey stars indicate Tonga earthquakes, the white triangles represent New Hebrides events. Arcuate regions (shaded to match the event symbols) are the zones on the CMB that contribute to PKP precursors. Parts of Australia are shown to the southwest. The black box is the region inferred to show very strong scattering.

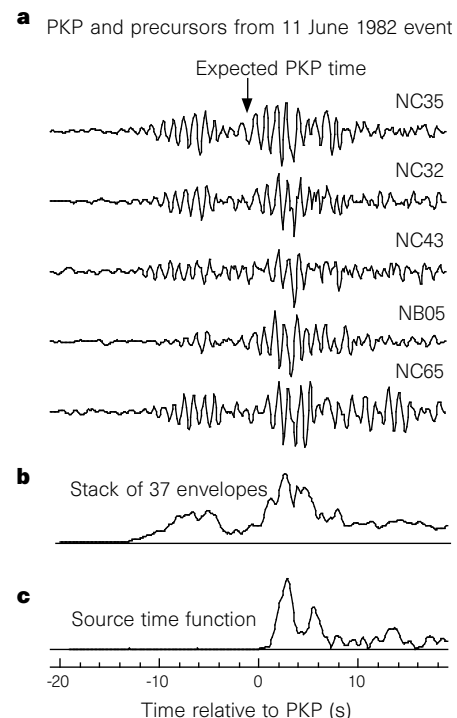


Figure 3 PKP precursor observations. We processed the data as follows: first, the data was filtered in the bandpass from 0.7 to 2.0 Hz to minimize microtremor as well as high-frequency noise; second, we took an envelope of the seismograms, as the data were not suitable for a coherent stack; and third, we stacked the data on the expected slowness of PKP waves. **a**, Five typical filtered seismograms from an event that occurred 136.4° from NORSAR. **b**, Slant stack of the envelopes of the filtered seismograms. **c**, Stack of P waves from California¹⁷ to show earthquake duration.

medium using Chernov's theory¹⁹ adapted to elastic media^{4,7}. An alternative possibility is that these precursors to PKP arise from coherent reflections from a dipping planar interface. However, below 0.5 Hz, the precursors in the distance range up to 140° are a factor of ten smaller than PKP. This frequency dependence verifies that the large arrivals we examine are indeed scattered waves.

Our preferred model for the origin of the scattered waves has 13% r.m.s. velocity variations with 10 km scale length in the lowermost 60 km of the mantle across a region at least spanning the box in Fig. 2. Models that extend further up into the mantle or further into the sampled regions do not reproduce the time dependence and thus give poorer data fits. Marginally better fits can be obtained by a model that allows heterogeneous structure only at the CMB. We do not favour this model as it requires still greater velocity variations. The required velocity perturbations scale inversely with the square root of the thickness: for example, the required perturbations scale from 23% for a 20-km-thick patch down to 6% for a patch spanning 250 km. Without lateral variation in the strength of scattering, the observed variations in the precursors that are not simply a function of range cannot be explained. The western edge of the patch limits

the predicted amplitudes in the New Hebrides trace. Figure 4 shows the predictions of the preferred model compared to scattering predicted from a model with 1% velocity variations across the entire mantle⁷. The global model predicts far less scattering than is observed. Strong scattering uniformly distributed near the CMB also does not fit the data well.

Global surveys of PKP precursors have failed to identify regions that scatter as strongly as the region we discuss here. A survey of the data set from a recent publication⁷ showed no other paths with PKP precursors larger than 20% of the amplitude of PKP for distances less than 137°. A previous study¹⁶ made the same observation. In contrast, our paths from northern Tonga to NORSAR shows precursors with comparable amplitude to PKP in the range from 136° to 138°, as shown in Fig. 4.

Our initial survey indicates that recordings of PKP precursors that sample low-velocity lower mantle other than the southwest Pacific do not show nearly such large precursor amplitudes. To investigate the CMB under Iceland, we have examined records from NORSAR of French nuclear tests in the Tuamotu archipelago. Their receiver-side scattering zone includes the region below Iceland. These data, mentioned above in reference to their source-side scattering patch, have noticeable but minor precursors. Beneath Africa, the most prominent low-velocity feature in the lower mantle^{8,20}, a spot-check of recordings in Tanzania of Tongan and Mexican earthquakes showed only normal PKP precursors (J. Ritsema, personal communication).

The region that scatters PKP waves lies in the 'equatorial Pacific plume group' (ref. 8) structure which is distinguished by low S-wave velocities in the lowermost several hundred kilometres of the mantle. P-wave tomography of this region²¹ also shows lower than average velocities. The low velocities imply hotter than average mantle. In fact, the 'equatorial Pacific plume group' structure that we sample and the 'great African plume' (ref. 8) structure under Africa are probably the two hottest spots at the base of the mantle. However, the S- and P-wave speed anomalies are only 3% and 1%, respectively.

The region about 1,000 km to the southeast of our study area has been inferred to have layering at the base of the mantle with 10–20% reduction in the lowermost tens of kilometres (refs 9, 15), which has been interpreted as partially molten mantle¹⁰, although some argue that it may be difficult to melt the lowermost mantle²². The CMB in some other regions, in contrast, appears free of strong scattering²³.

The 13% r.m.s. amplitude of the P-wave velocity variations in our scattering region require the presence of partial melt¹⁰, as other possibilities that could explain weaker velocity heterogeneity at the base of the mantle do not work¹². In addition, the large lateral gradients in velocity are likely to be accompanied by significant lateral gradients in density. The simplest way to maintain such structure is vigorous convection within the thermal boundary layer at the base of the mantle. □

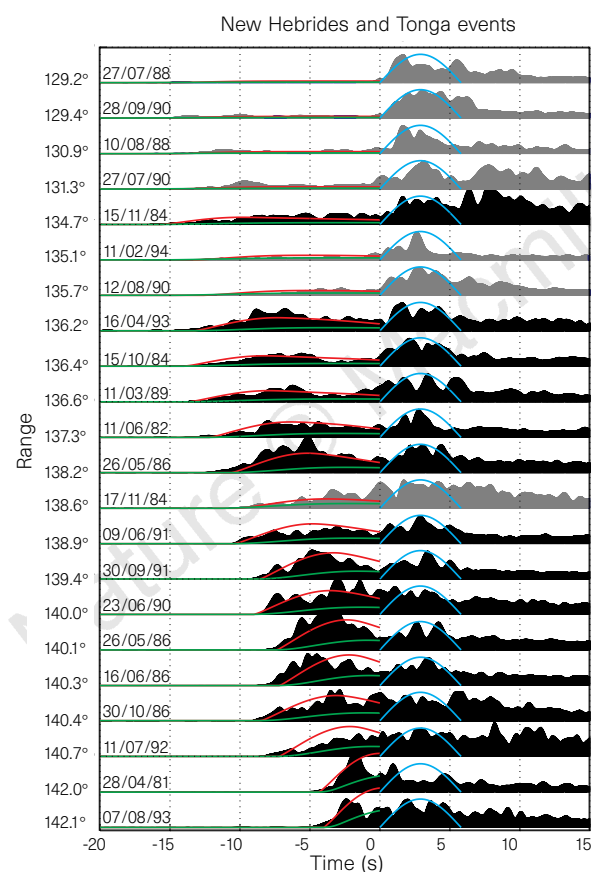


Figure 4 Data and synthetic curves. Stacks of envelopes of bandpassed seismograms for all events are arranged by distance; amplitudes are adjusted downwards by an amount determined by the average energy level of the noise⁷, normalized to unity and shifted slightly in time to align PKP. The Tongan records are in black. Dates of events are given as day/month/year. Two sets of calculated envelopes are shown to the left of zero time. The red lines show calculated PKP precursors for 13% r.m.s. velocity variations in the lowermost 60 km of the mantle in the box shown in Fig. 1. The green, lower-amplitude, lines show the calculation for 1% r.m.s. variation across the entire mantle, as determined by a global study⁷. The source-time function is represented by the blue curves to the right of zero time.

Received 2 June 1997; accepted 13 January 1998.

- Haddon, R. A. W. Corrugations on the CMB or transition layers between inner and outer cores? *Eos* **53**, 600 (1972).
- Cleary, J. R. & Haddon, R. A. W. Seismic wave scattering near the core-mantle boundary: a new interpretation of precursors to PKP. *Nature* **240**, 549–551 (1972).
- Doornbos, D. J. & Husebye, E. S. Array analysis of PKP phases and their precursors. *Phys. Earth Planet. Inter.* **5**, 387–399 (1972).
- Haddon, R. A. & Cleary, R. J. Evidence for scattering of seismic PKP waves near the core-mantle boundary. *Phys. Earth Planet. Inter.* **8**, 211–234 (1974).
- King, D. W., Haddon, R. A. W. & Cleary, J. R. Array analysis of precursors to PKIPK in the distance range 129° to 142°. *Geophys. J. R. Astron. Soc.* **37**, 157–173 (1974).
- Husebye, E. S., King, D. W. & Haddon, R. A. W. Precursors to PKIPK and seismic wave scattering near core-mantle boundary. *J. Geophys. Res.* **81**, 1870–1882 (1976).
- Hedlin, M. A. H., Shearer, P. M. & Earle, P. S. Seismic evidence for small-scale heterogeneity throughout the Earth's mantle. *Nature* **387**, 145–150 (1997).
- Su, W. J., Woodward, R. L. & Dziewonski, A. M. Degree-12 model of shear velocity heterogeneity in the mantle. *J. Geophys. Res.* **99**, 6945–6980 (1994).
- Garnero, E. J. & Helmberger, D. V. Seismic detection of a thin laterally varying boundary layer at the base of the mantle beneath the central-Pacific. *Geophys. Res. Lett.* **23**, 977–980 (1996).
- Williams, Q. & Garnero, E. J. Seismic evidence for partial melt at the base of the Earth's mantle. *Science* **273**, 1528–1530 (1996).

11. Jeanloz, R. & Richter, F. M. Convection, composition, and the thermal state of the lower mantle. *J. Geophys. Res.* **84**, 5497–5504 (1979).
12. Manga, M. & Jeanloz, R. Implications of metal-bearing chemical boundary layer in D'' for mantle dynamics. *Geophys. Res. Lett.* **23**, 3091–3094 (1996).
13. Lay, T. Structure of the core-mantle transition zone; a chemical and thermal boundary layer. *Eos* **70**, 49–49 (1989).
14. Loper, D. E. & Lay, T. The core-mantle boundary region. *J. Geophys. Res.* **100**, 6397–6421 (1995).
15. Mori, J. & Helmberger, D. V. Localized boundary layer below the mid-Pacific velocity anomaly identified from a PcP precursor. *J. Geophys. Res.* **100**, 20359–20365 (1995).
16. Bataille, K. & Flatté, S. M. Inhomogeneities near the core-mantle boundary inferred from short-period scattered PKP waves recorded at the global digital seismograph network. *J. Geophys. Res.* **93**, 15057–15064 (1988).
17. Vidale, J. E. & Houston, H. The depth dependence of earthquake duration and implications for rupture mechanisms. *Nature* **365**, 45–47 (1993).
18. Doornbos, D. J. & Vlaar, N. J. Regions of seismic wave scattering in the Earth's mantle and precursors to PKP. *Nature* **243**, 58–61 (1973).
19. Chernov, L. A. *Wave Propagation in a Random Medium* (trans. Silverman, R. A.) (McGraw-Hill, New York, 1960).
20. Grand, S. P. Mantle shear structure beneath the Americas and surrounding oceans. *J. Geophys. Res.* **99**, 11591–11622 (1994).
21. van der Hilst, R., Widiyantoro, S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).
22. Zerr, A., Serghiou, G. & Boehler, R. Melting of CaSiO₃ perovskite to 430 kbar and first in situ measurements of lower mantle eutectic temperatures. *Geophys. Res. Lett.* **24**, 909–912 (1997).
23. Vidale, J. E. & Benz, H. M. A sharp and flat section of the core-mantle boundary. *Nature* **359**, 627–629 (1992).

Acknowledgements. We thank H.-C. Nataf for a review; Q. Williams, S. Grand, E. Garnero, P. Shearer and D. Helmberger for discussions; J. Fyen and J. Torstveit for supplying the NORSAR data; and J. Ritsema for evaluating the PKP precursors from Tanzanian stations. This work was supported by the US NSF.

Correspondence and requests for materials should be addressed to J.E.V. (e-mail: vidale@ucla.edu).

The breeding structure of a tropical keystone plant resource

John D. Nason*, E. Allen Herre† & J. L. Hamrick‡

* Department of Biological Sciences, University of Iowa, 312 Chemistry Building, Iowa City, Iowa 52242, USA

† Smithsonian Tropical Research Institute, Apartado 2072, Balboa, Ancon, Republic of Panama

‡ Departments of Botany and Genetics, University of Georgia, Athens, Georgia 30602, USA

Despite the recognized importance of maintaining viable populations of keystone plant resources in tropical wildlife parks and forested preserves, the critical question of what constitutes effective breeding units of these species has not been directly addressed. Here we use paternity analysis techniques to reconstruct the genotypes of pollen donor trees and to estimate pollen dispersal distances and breeding population size parameters for Panamanian populations of seven species of monoecious strangler figs (*Ficus*, Moraceae), a particularly widespread and influential group of keystone producers^{1–3}. Despite the minute size (1–2 mm) and short lifespan (2–3 d) of the species-specific wasp pollinators (Agaonidae, Chalcidoidea), pollen dispersal was estimated to occur routinely over distances of 5.8–14.2 km between widely spaced host trees. As a result of such extensive pollen movement, breeding units of figs comprise hundreds of intermating individuals distributed over areas of 106–632 km², an order of magnitude larger than has been documented for any other plant species. Moreover, these results should be generalizable to the 350 or so monoecious fig species that share this pollination system⁴. The large areal extent of breeding units of these keystone plant resources has important implications for our understanding of both the evolution of tropical biodiversity and its maintenance by applied conservation efforts.

Like a keystone supporting an archway, keystone tropical plant resources fruit all year round and so support a broad spectrum of vertebrate frugivores during times of food scarcity. Figs are considered to be the pre-eminent group of keystone plant resources in

southeast Asia and in the Neotropics owing to their heavy fruit production and generally aseasonal patterns of reproduction^{1–3}. Given their disproportionately strong influence on species assemblages at many trophic levels, a failure to maintain viable fig populations in forest preserves is expected to result in a cascade of subsequent extinction events². Information guiding the establishment of reserve areas sufficient for the long-term preservation of fig populations is therefore of vital importance to the management of tropical biodiversity, particularly as large areas of continuous tropical forest become increasingly fragmented by human activity.

We have examined spatial patterns of effective fig wasp and pollen dispersal by using paternity analysis techniques to reconstruct precisely the isozyme genotypes of pollen donors fertilizing the fruit of individual fig trees. The research was conducted on populations of seven monoecious fig species (subgenus *Urostigma* sect. *Americana*) and their wasps (genus *Pegoscapus*) occurring in central Panamá. Adults of the fig species studied, *F. citrifolia*, *F. dugandii*, *F. nymphiifolia*, *F. obtusifolia*, *F. perforata*, *F. pertusa* and *F. popenoei*, occur at average population densities of less than ten individuals per square kilometre on Barro Colorado Island (BCI; C. Handley and E. Kalko, unpublished census data)^{5,6}, a large, 15-km² nature preserve located within the study area. Typical of monoecious figs, these species are obligately outcrossing and produce large fruit crops with reproduction synchronous within individual trees but asynchronous between trees^{6,7}, so that co-flowering individuals are often located substantial distances apart.

The power to reconstruct pollen donor genotypes was facilitated by high levels of assayable genetic variation and access to full-sibling

Table 1 Genetic diversity and probabilities of paternity exclusion

<i>Ficus</i> sp.	Genetic diversity*	Possible isozyme genotypes	Paternity exclusion probability	
			Single offspring	Full-sib progeny array†
<i>F. citrifolia</i>	0.258	15,360	0.769	0.997
<i>F. dugandii</i>	0.235	73,728	0.830	>0.999
<i>F. nymphiifolia</i>	0.169	15,360	0.720	0.995
<i>F. obtusifolia</i>	0.231	13,271,040	0.887	>0.999
<i>F. perforata</i>	0.227	331,776	0.912	>0.999
<i>F. pertusa</i>	0.258	31,104	0.869	>0.999
<i>F. popenoei</i>	0.223	20,736	0.821	>0.999

The precision with which pollen donor genotypes was reconstructed was facilitated by highly variable genetic marker loci and access to single foundress, singly sired fruit. Based on observed genetic diversity and allelic variation, thousands of multilocus allozyme genotypes are possible for each species. As a result, the probability of paternity exclusion, a measure of the power with which the genotype of the actual paternal parent can be distinguished from other putative fathers, is very high (>0.995). This probability is greatly enhanced in the fig species studied because full-sib progeny arrays, as opposed to single offspring, can be used to infer the father's multilocus genotype.

* Calculated as average expected heterozygosity over polymorphic and monomorphic loci.

† Calculated as one minus the probability of genotypic identity.

Table 2 Pollen donor diversity and breeding population size

<i>Ficus</i> species	Trees sampled	No. of fruit per tree	Pollen parents		Breeding unit size	
			$d_{\min}$	$\bar{d}^*$	$\hat{N}_{d_{\min}}$	$\hat{N}_{\bar{d}^*}$
<i>F. citrifolia</i>	2	35.0	13.0	20.5 (2.7)	182.5	287.8 (38.3)
<i>F. dugandii</i>	1	15.0	11.0	18.0 (3.7)	154.4	252.7 (51.9)
<i>F. nymphiifolia</i>	2	28.5	10.0	23.0 (5.5)	140.4	322.9 (77.0)
<i>F. obtusifolia</i>	7	26.2	17.3	54.3 (5.6)	242.7	762.1 (79.0)
<i>F. perforata</i>	2	20.5	6.0	11.0 (2.7)	84.2	154.2 (37.2)
<i>F. pertusa</i>	2	16.0	10.0	16.0 (2.5)	140.9	224.6 (34.3)
<i>F. popenoei</i>	3	14.0	10.7	28.0 (5.3)	147.7	393.1 (74.0)

The number of functionally male, staminate phase trees in the breeding population surrounding a given maternal tree is indicated by the total number of different pollen parent genotypes represented in its fruit crop ($\bar{d}$) and was estimated in two ways, from the observed number of distinguishable paternal genotypes in a sample of singly sired fruit ($d_{\min}$, a lower bound), and, more accurately, from the frequency distribution of different paternal genotypes revealed in such a sample ($\bar{d}$, estimated according to ref. 22). Given the reproductive phenologies of the fig populations studied, the observed patterns of paternity indicate that, despite low population densities, breeding units centred about maternal fig trees consist of hundreds of intermating individuals.

* Standard error in parentheses.

progeny arrays obtained from singly sired fruits (Table 1). Levels of allozyme genetic diversity for the fig species studied (mean 0.229) are significantly higher than population level estimates for other very low-density tropical trees (mean 0.124; ref. 8) and for tropical and temperate woody species in general (mean 0.125 and 0.146, respectively⁸). Given this level of genetic diversity at 7–15 polymorphic loci, analysis of 12–18 full-sibling progeny per fig fruit (syconia) permits very precise reconstruction of the multilocus allozyme genotype of the pollen parent. In the fig species studied, dispersing female fig wasps become trapped in and do not leave the receptive, female-phase syconia post-oviposition and pollination. Thus, full-sibling progeny arrays can be identified before genetic analysis from syconia found to contain the remains of a single wasp foundress⁹.

Paternity analysis revealed that maternal trees of each species were fertilized by numerous, genetically distinguishable pollen donors and that the numbers of pollen donors detected would have continued to rise with additional sampling (Table 2). From information on flowering phenologies, the inferred patterns of paternity indicate that breeding units of the study species consist of 150 to more than 750 individuals (Table 2), a substantially larger number of trees than are known to occur on BCI. For the three species that have been censused extensively on BCI (*F. dugandii*, *F. obtusifolia* and *F. popenoei*), breeding units are estimated to occupy areas of 106–632 km², with fig wasps routinely dispersing pollen over distances of 5.8–14.2 km to receptive flowering trees (Table 3).

The breeding unit areas and intrapopulation pollen dispersal distances described here from monoecious figs are considerably larger than those described for any other animal- or wind-pollinated, tropical or temperate tree species (Table 3). Furthermore, because strangler figs of the section *Americana* are close to Old World *Urostigma* species evolutionarily and in terms of growth form and the highly stereotyped system of pollination by small, species-specific wasps¹⁰, these breeding structures should be generalizable to hundreds of fig species worldwide. Although the mechanism of dispersal over such long distances is poorly understood, one plausible hypothesis is that passive, wind-mediated transport may serve a primary role. Indeed, the direction and velocity of wind currents influence patterns of wasp arrival to and departure from

fruiting fig trees^{11,12} and may account for the occasional long-distance movement of fig wasps over distances of several to tens of kilometres^{13–15}. Although a variety of small insects (including Diptera, Homoptera and Hymenoptera) are known to be wind-dispersed, it has not been demonstrated before that such movement is a natural and routine part of a pollinator's dispersal biology for fig wasps or any other pollinator species.

The capacity for extensive dispersal by fig wasps and the diffuse breeding structure of fig populations has important consequences for the management of diversity in tropical landscapes fragmented by human activity. Because animals are the primary agents of pollen dispersal in tropical trees, the negative reproductive and genetic effects of habitat fragmentation will be most pronounced when vector dispersal is curtailed by fragmentation. Under these conditions fragments become isolated populations of small effective population size, increasingly subject to stochastic demographic and genetic processes, including inbreeding and loss of adaptive genetic variation. The breeding structure of fig populations, as with any plant species, will change when habitat fragmentation results in the loss of reproductive individuals¹⁶. But in contrast to most species, the fig and fig wasp mutualism can apparently form extensive metapopulations in fragmented landscapes, with subpopulations linked both reproductively and genetically by the exceptional dispersal capabilities of the wasp pollinators. Moreover, although frugivore diversity is known to decline as a result of habitat disturbance and disruption of population and community structures¹⁷, isolated fig trees occurring in small forest fragments can contribute to the survival of figs, and vertebrate assemblages, in large forest remnants and reserves. Indeed, on the basis of simulation study estimates of the fig population sizes needed to maintain an uninterrupted temporal sequence of oviposition sites for fig wasps^{18–20}, populations of most of the 12 strangler fig species on BCI consist of too few individuals to keep the species-specific pollinators from local extinction. If isolated from other conspecifics, these populations should exhibit pollinator-limited reproductive failure. Reproductive success is routine on BCI^{6,11}, however, and, as indicated by patterns of paternity (Tables 2 and 3), may be largely sustained by interactions with conspecifics located in surrounding forests.

Table 3 Breeding unit parameters estimated for tropical and temperate tree species

Species	Pollen vector	Density (ha ⁻¹)	Adults	Breeding unit parameters	
				Area (km ²)	Radius (km)
Monoecious figs*					
<i>Ficus dugandii</i>	Fig wasp	0.004	252.7 (51.9)†	631.7 (129.9)†	14.2 (10.9–16.9)‡
<i>Ficus obtusifolia</i>	Fig wasp	0.072	762.1 (79.0)†	105.9 (11.0)†	5.8 (5.2–6.4)‡
<i>Ficus popenoei</i>	Fig wasp	0.013	393.1 (74.0)†	294.8 (55.5)†	9.7 (7.7–11.4)‡
Tropical trees					
<i>Astrocaryum mexicanum</i> ²⁴	Beetle	1364	1542	0.011	0.060
<i>Calophyllum longifolium</i> ²⁵	Small insect	0.28	35	1.241	0.629
<i>Cordia alliodora</i> ²⁶	Small insect	20.9	520	0.249	0.282
<i>Pithecellobium elegans</i> ²⁷	Hawkmoth	0.88	45	0.636	0.450
<i>Platypodium elegans</i> ²⁸	Small bees	0.78	68	0.866	0.525
<i>Spondias mombin</i> ²⁵	Small insect	0.33	6	0.196	0.250
<i>Turpinia occidentalis</i> ²⁵	Unknown	1.27	5	0.040	0.113
Temperate trees					
<i>Cedrus atlantica</i> ²⁹	Wind	61.7	934	0.151	0.220
<i>Fraxinus americana</i> ²⁹	Wind	24.7	20	0.008	0.050
<i>F. pennsylvanica</i> ²⁹	Wind	61.7	49	0.008	0.050
<i>Pinus cembroides</i> ²⁹	Wind	61.7	49	0.008	0.050
<i>Pinus radiata</i> ³⁰	Wind	2.5	40	0.159	0.225
<i>Populus deltoides</i> ²⁹	Wind	24.7	25,942	10.507	1.829
<i>Pseudotsuga menziesii</i> ³¹	Wind	25	196	0.078	0.158
<i>Pseudotsuga taxifolia</i> ²⁹	Wind	128.4	121	0.010	0.055
<i>Ulmus americana</i> ²⁹	Wind	24.7	31,389	12.714	2.012

The spatial dimensions and numbers of adults characterizing breeding units of three monoecious figs are compared to those of other animal-pollinated tropical and wind-pollinated temperate tree species. For the non-*Ficus* species, the breeding unit corresponds to Levin's²² paternity-pool concept and represents the circular area about a female plant within which 99% of potential mates are expected to occur. Estimates are averages over females and, because breeding units of herbs are much smaller²³, are presented for tree species only.

* Breeding unit parameters estimated based on N_b (Table 2).

† Standard errors in parentheses.

‡ 95% confidence limits in parentheses (standard errors are asymmetrical).

In summary, breeding units of the monoecious fig species studied are distributed over a very large area, many times larger than BCI itself (Table 3). As a result, the ability of this reserve to maintain viable populations of these keystone plant resources and, by association, many other plant and animal species, is strongly linked to the management of surrounding forested areas. Moreover, patterns of mating indicate that populations of figs may remain reproductively viable under conditions of insular, and presumably terrestrial, habitat fragmentation, so long as the number of adult trees within the dispersal radius of female wasps does not fall below the minimum critical size needed to support populations of the pollinator. The breeding structure and pollen dispersal dynamics of other keystone plant species will require further investigation before their responses to habitat fragmentation can be accurately predicted. □

Methods

Breeding unit size. Donor number is related to the size of the breeding unit (N individuals) as $d = Np$, where d is the number of different pollen parent genotypes represented in a maternal tree's fruit crop and p is the expected proportion of trees in staminate phase at any given time. Assuming, conservatively, that individual trees reproduce asynchronously twice per year⁶ and that staminate and pistillate flowering phases each last seven days²⁰, p is calculated to be 0.0712. Although these estimates could be effected by mosaicism²¹ (the fusion of genetically distinct individuals), its frequency in the species examined here is low.

Breeding unit area and radii. These parameters were calculated from paternity-analysis-based estimates of breeding unit size (N_B ; Table 2) and the censused densities of adult, reproductively mature trees over 15 km² BCI. Based on the long-term censuses of C. Handley and E. Kalko, 6, 108 and 20 adults of *F. dugandii*, *F. obtusifolia* and *F. popenoei*, respectively, are known to occur on BCI. Because these species exhibit little spatial aggregation over the area censused, these densities are assumed to be representative of forested areas surrounding BCI, a conservative assumption given that approximately one-third of the area within 10 km of BCI is occupied by Lake Gatun where figs are absent. Breeding unit radii estimate the distances pollen-bearing, female fig wasps routinely disperse in search of receptive host trees. Although actual breeding populations of figs may deviate substantially from the assumed circular distribution, alternative structures (elliptical, for example) increase estimated wasp dispersal distances.

Received 24 April; accepted 21 October 1997.

- Leighton, M. & Leighton, D. R. in *Tropical Rain Forest: Ecology and Management* (eds Sutton, S. L., Whitmore, T. C. & Chadwick, A. C.) 181–196 (Blackwell, Oxford, 1983).
- Terborgh, J. in *Conservation Biology: The Science of Scarceness and Diversity* (ed. Soulé, M. E.) 330–344 (Sinauer, Sunderland, MA, 1986).
- Lambert, F. R. & Marshall, A. G. Keystone characteristics of bird dispersed *Ficus* in a Malaysian lowland rain forest. *J. Ecol.* **79**, 793–809 (1991).
- Wiebes, J. T. Agaonidae (Hymenoptera, Chalcidoidea) and *Ficus* (Moraceae): fig wasps and their figs, xv (Meso-American Pegoscapus). *Proc. K. Ned. Akad. Wet.* **98**, 167–183 (1995).
- Todzia, C. Growth habits, host tree species, and density of hemiepiphytes on Barro Colorado Island, Panama. *Biotropica* **18**, 22–27 (1986).
- Windsor, D. M., Morrison, D. W., Estribi, M. A. & Leon, B. D. Phenology of fruit and leaf production by 'strangler' figs on Barro Colorado Island, Panamá. *Experientia* **45**, 647–653 (1989).
- Janzen, D. H. How to be a fig. *Annu. Rev. Ecol. Syst.* **10**, 13–51 (1979).
- Hamrick, J. L. in *Proceedings of the International Symposium on Genetic Conservation and Production of Tropical Forest Tree Seed* (eds Drysdale, R. M., John, S. E. T. & Yapa, A. C.) 1–9 (Asean-Canada Forest Tree Seed Centre, 1994).
- Herre, E. A. Coevolution of reproductive characteristics in 12 species of New World figs and their pollinator wasps. *Experientia* **45**, 637–647 (1989).
- Herre, E. A. *et al.* Molecular phylogenies of figs and their pollinating wasps. *J. Biogeogr.* **23**, 521–530 (1996).
- Ware, A. B. & Compton, S. G. Dispersal of adult female fig wasps. 1. Arrivals and departures. *Entomol. Exp. Appl.* **73**, 221–229 (1994).
- Ware, A. B. & Compton, S. G. Dispersal of adult female fig wasps. 2. Movements between trees. *Entomol. Exp. Appl.* **73**, 231–238 (1994).
- Compton, S. G. A collapse of host specificity in some South African fig wasps. *S. African J. Sci.* **86**, 39–40 (1990).
- Compton, S. G., Ross, S. J. & Thornton, I. W. B. Pollinator limitation of fig tree reproduction on the island of Anak Krakatau (Indonesia). *Biotropica* **26**, 180–186 (1994).
- McKey, D. Population biology of figs: applications for conservation. *Experientia* **45**, 661–673 (1989).
- Nason, J. D., Aldrich, P. R. & Hamrick, J. L. in *Tropical Forest Remnants: Ecology, Management and Conservation of Fragmented Communities* (eds Laurance, W. F. & Bierregaard, R. O. Jr) 304–320 (University of Chicago Press, Chicago, 1997).
- Bierregaard, R. O. Jr, Lovejoy, T. E., Kapos, V., Santos, A. A. D. & Hutchings, R. W. The biological dynamics of tropical rainforest fragments: A prospective comparison of fragments and continuous forest. *Bioscience* **42**, 859–866 (1992).

- Kjellberg, F. & Maurice, S. Seasonality in the reproductive phenology of *Ficus*: its evolution and consequences. *Experientia* **45**, 653–660 (1989).
- Bronstein, J. L., Gouyon, P.-H., Gliddon, C., Kjellberg, F. & Michaloud, G. The ecological consequences of flowering asynchrony in monoecious figs: a simulation study. *Ecology* **71**, 2145–2156 (1990).
- Anstett, M. C., Hossaert-McKey, M. & McKey, D. Modelling the persistence of small populations of strongly interdependent species of figs and fig wasps. *Conserv. Biol.* **11**, 204–213 (1997).
- Thomson, J. D., Herre, E. A., Hamrick, J. L. & Stone, J. L. Genetic mosaics in strangler fig trees: implications for tropical conservation. *Science* **254**, 1214–1216 (1991).
- Burnham, K. P. & Overton, W. S. Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**, 927–936 (1979).
- Levin, D. A. The paternity pools of plants. *Am. Nat.* **132**, 309–317 (1988).
- Eguarte, L. E., Pérez-Nasser, N. & Pinero, D. Genetic structure, outcrossing rate and heterosis in *Astrocaryum mexicanum* (tropical palm): implications for evolution and conservation. *Heredity* **69**, 217–228 (1992).
- Stacy, E. A. *et al.* Pollen dispersal in low density populations of three neotropical tree species. *Am. Nat.* **148**, 275–298 (1996).
- Boshier, D. H., Chase, M. R. & Bawa, K. S. Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree. 3. Gene flow, neighborhood, and population structure. *Am. J. Bot.* **82**, 484–490 (1995).
- Chase, M. R., Moller, C., Kessell, R. & Bawa, K. S. Distant gene flow in tropical trees. *Nature* **383**, 398–399 (1996).
- Nason, J. D. & Hamrick, J. L. Reproductive and genetic consequences of forest fragmentation: two case studies of Neotropical canopy trees. *J. Hered.* **88**, 264–276 (1997).
- Wright, J. Pollen-dispersal studies: some practical applications. *J. Forest.* **51**, 114–118 (1953).
- Bannister, M. H. in *The Genetic of Colonizing Species* (eds Baker, H. G. & Stebbins, G. L.) 353–372 (Academic, New York, 1965).
- Adams, W. T. Gene dispersal within forest tree populations. *New Forests* **6**, 217–240 (1992).

Acknowledgements. We thank P. Aldrich, N. Ellstrand, T. Fleming, M. J. Godt, A. Graffen, K. Harms, M. Harris, E. Kalko, E. Leigh, R. May, K. Milton, S. Rand, J. Thomson and N. Waser for discussion. This work was supported by a grant from the NSF and was greatly aided by the Smithsonian Tropical Research Institute and its facilities on BCI.

Correspondence and requests for materials should be addressed to J.D.N. (e-mail: john-nason@uiowa.edu).

Female–female cooperation in polygynous oystercatchers

Dik Heg* & Rob van Treuren*†

* Zoological Laboratory and † Department of Genetics, University of Groningen, Kerklaan 30, PO Box 14, 9750 AA, Haren, The Netherlands

Waders (Charadrii) provide biologists with an astonishing variety of mating systems to study¹. Male and female birds establish breeding units in which behaviour varies from monogamy, polygyny, polyandry, double clutching, lekking and serial monogamy to sex role reversal, and many mixed mating systems exist¹. This diversity is currently explained by the costs and benefits of males and females either cooperating or defecting during breeding attempts^{2,3}. The oystercatcher (*Haematopus ostralegus*) is a typically monogamous species: removal experiments show that both parents are needed to raise chicks to fledgings^{4–6}. However, occasional polygyny has also been reported⁷. Here we describe polygynous oystercatcher trios and the reproductive consequences of such polygyny. There is a 'classical' form of polygyny (two female territories within the male territory), but oystercatchers also show a remarkable variant, accompanied by female–female cooperation, female–female copulations and joint nesting.

Polygyny was extremely rare in the main study population. Only $1.85 \pm 0.52\%$ (mean of 14 years of study $\pm$ s.e.m.) of the breeding males and $2.70 \pm 0.67\%$ of the breeding females were polygynous. Polygyny usually developed as a result of 'failed' female usurpations⁸ ending in a draw; in other words, females postponed fighting (this occurred in 89% of cases; $n = 19$; usurpations were initiated by 11 non-breeders and 6 neighbouring breeders). More rarely, polygyny was due to a male breeder extending his territory to include the territory of a widowed female (5% of cases) or a male non-breeder joining two female widows which were polygynously mated to a male in the previous year (5% of cases). In 57% of the trios ($n = 60$ females), polygynous females remained highly aggressive towards each other within the male's territory, but defended their half of the

territory cooperatively. This form of polygyny, which we name aggressive polygyny (AP), is widespread in birds³. In the other 43% of polygynous trios, the females stopped mutual aggression and started to preen close together and act cooperatively (37% of the total number of trios displayed this behaviour in the same breeding season, and 7% in their second breeding season). This form of polygyny, which we name cooperative polygyny (CP), has previously been incompletely described (see ref. 7 and references therein). In these cases, the two females and the male share, attend and defend one territory and one nest (Table 1). The two females also copulate regularly with each other and with the male (Table 1). Most strikingly, the female–female copulations appeared behaviourally indistinguishable from male–female copulations. The females irregularly adopted the male or the female copulation role⁹, sometimes in rapid succession. Females often aborted lesbian copulation sequences during initiation or before they achieved cloacal contact⁹ (53.1% of cases, $n = 32$), unlike in both male–female copulations within polygynous trios (26.3% of copulation sequences were aborted, $n = 165$) and male–female copulations in monogamous

pairs (33.3%, $n = 5711$; these are data from 1988–1995, likelihood ratio (LR) = 5.28, degrees of freedom (d.f.) = 2, $P = 0.07$; heterosexual compared with lesbian copulations: LR = 2.58, d.f. = 1, $P = 0.11$). It has been argued that mates in monogamous pairs use frequent copulations to signal their cooperative pair bond⁹. Similarly, CP females might signal and support their cooperative bond by frequent female–female copulations. The bonobo has been reported to behave homosexually in a similar manner to cooperative oystercatcher females, apparently to appease intrasexual conflicts (see ref. 10 and references therein). The pukeko (a group-living bird) behaves similarly; in this case, the behaviour probably relates to both intersexual and intrasexual competition¹¹.

Why do oystercatcher females establish cooperative bonds with other females? Kin selection¹² might facilitate the occurrence of cooperative polygyny. To test this, we compared the genotypes of the two females within a polygynous breeding trio (Fig. 1). None of the coefficient of relatedness (r_{xy}) values between these females fall within the right-hand 5% of the distribution. This indicates that, compared with randomly chosen pairs, these females were not

Table 1 Behaviour of cooperative and aggressive polygynous trios compared with behaviour of monogamous pairs

Mating system	Joint nesting*	Males incubating*	Male–female copulation frequency†	Female–female copulation frequency†	% Male–female cooperative territory defence†	% Female–female cooperative territory defence†	% Inter-female aggression†	% Female brooding†	% Male brooding†
Cooperative polygyny									
Female 1	95 (19)	100 (19)	0.26 ± 0.06 (40)	0.15 ± 0.04 (54)	18.2 ± 4.5 (4)	10.4 ± 3.5 (8)	0.00 ± 0.00 (8)	22.7 ± 8.4 (18)	19.8 ± 6.4 (18)
Female 2	95 (19)	100 (19)	0.18 ± 0.04 (40)		9.2 ± 2.9 (4)			57.5 ± 8.9 (18)	
Both females					40.3 ± 13.4 (4)				
Aggressive polygyny									
Female 1	0 (18)	94 (18)	0.24 ± 0.06 (34)	0.00 ± 0.00 (43)	12.9 ± 2.5 (5)	0.0 ± 0.0 (5)	31.5 ± 15.1 (5)	28.1 ± 9.4 (9)	54.3 ± 10.9 (9)
Female 2	0 (20)	65 (20)	0.17 ± 0.04 (34)		6.0 ± 3.3 (5)			17.6 ± 12.1 (9)	
Monogamy									
Female	–	100 (1,404)	0.28 ± 0.03 (173)	–	44.3 ± 4.2 (42)	–	–	51.2 ± 4.8 (57)	43.3 ± 4.6 (57)

* Columns refer to all first clutches of polygynous trios, 1983–1997. Sample sizes are given in parentheses. In some cases no clutch was found for one or two females. Joint nesting: percentage of clutches laid in one nest that were incubated by the females from one polygynous trio, excluding cases in which no clutch was found from one or both the females. Males incubating: percentage of clutches incubated by the males. The probability of male help during brooding depended on the mating system (categorical logistic regression: G-test datum ($G = 60.4$, d.f. = 2, $P < 0.00001$)). Within polygynous trios, male help depended on the type of polygyny ($G = 12.3$, d.f. = 1, $P < 0.001$) and males helped their second female less ($G = 5.5$, d.f. = 1, $P < 0.02$). † 1995 results. Sample sizes are in parentheses. Copulation frequencies: the mean number of successful copulations per hour per observation in the prelaying phase. Territory defence and aggression: mean percentages of cooperation between the male and female(s) against conspecifics and, in the case of polygyny, percentages of cooperation between female 1 and female 2 and aggression between female 1 and female 2. Brooding: mean percentage brooding of the total time spent brooding by all individuals within the pair or the trio per observation. Copulation frequency of monogamous females compared to polygynous females: ANOVA: $P > 0.7$. Repeated measurements ANOVA of polygynous females comparing the effects of female (1 or 2) and type of polygyny (CP or AP): female $F = 2.87$, $P = 0.10$; type $F = 0.09$, $P = 0.77$; interaction $F = 0.01$, $P = 0.93$. Male–female cooperative territory defence, U tests: CP trios cooperative defence by male and two females versus monogamous pairs ($U = 49$, $P = 0.33$). Cooperation with the male in polygynous trios, comparing CP with AP trios, U tests: CP female 2 versus AP female 1 ($U = 6$, $P = 0.33$); CP female 2 versus AP female 2 ($U = 7$, $P = 0.46$). Males within polygynous trios assisted their second female less; CP and AP trios combined: Wilcoxon's $Z = -2.03$, $P = 0.04$. Brooding: monogamous males and females compared with polygynous males and females; all differences $> 10\%$ are significantly different, except for monogamous females versus AP females 1 (U tests). Differences within polygynous trios: only within CP trios, female 2 took a significantly larger share of incubation compared with both the male and female 1 (Wilcoxon's tests, both $P < 0.05$, $n = 18$).

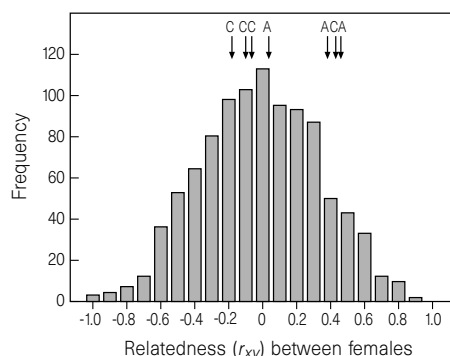


Figure 1 Frequency distribution of the degree of relatedness (r_{xy}) for 1,000 randomly generated pairs of female oystercatchers on the island of Schiermonnikoog. Pairwise r_{xy} values between the females from seven polygynous breeding groups are shown by arrows above corresponding relatedness classes, where 'A' denotes aggressive polygynous females and 'C' denotes cooperative polygynous females. The exact values for C are -0.23 , -0.13 , -0.07 and 0.38 , and for A are 0.04 , 0.35 and 0.41 .

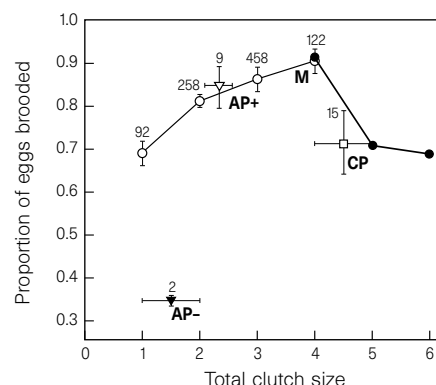


Figure 2 Mean proportion of brooded eggs in the nest for monogamous pairs (M, open circles; clutch sizes are connected with lines), aggressive polygynous trios with male help (AP+, open triangle) and without male help (AP-, filled triangle), cooperative polygynous trios (CP, open square; clutches of both females were combined, because they were brooded in one nest) and experimental monogamous nests containing copper eggs (filled circles). See Methods for statistical analysis. Sample sizes (number of nests) are indicated in the graph.

genetically related to a higher degree. Therefore, kin selection is not likely to underlie the occurrence of polygyny.

Another explanation might be that females benefit more from helping each other than from breeding monogamously; this phenomenon is called mutualism. However, polygynous females produced fewer fledglings, and their survival rate was not higher than that of monogamous females (Table 2). Both CP and AP females failed to hatch many eggs. In AP trios the females competed heavily with each other, including during incubation. As the male helped only the female that was first to lay eggs or had to care for both unguarded nests (in six trios, the mean difference in laying date of the two females was 16.04 ± 4.3 days, range 2–28 days; incubation lasts 28 days), this frequently led to predation of the unguarded nests¹³. But why did so many CP clutches fail to hatch? Incomplete brooding of the large, joint CP clutches might account for the high number of hatching failures (Table 2). Only one trio member brooded the combined clutches of both females, although all three members brooded the clutches alternately. As CP females tended to synchronize their egg-laying behaviour (the mean difference in laying date of the two females was 1.2 ± 0.7 days, range 0–4 days, $n = 6$) and laid their eggs in one nest, this joint nest contained up to seven eggs during the largest part of the incubation period (mean = 4.6 ± 0.4 eggs, $n = 23$, frequency distribution 1–7 eggs: 1, 2, 4, 3, 4, 6, 3). Monogamous females laid up to four eggs (mean number of eggs laid in the first clutches = 2.6 ± 0.02 , $n = 2240$; some monogamous females laid a fifth egg to compensate for egg loss during the laying period, 0.5% of the clutches). We hypothesized that four eggs might be the maximum number brooded effectively at a time by one adult, given that the eggs and the brood patch have a certain size. We tested this hypothesis by analysing brooding on CP clutches containing a total of five to seven eggs (the combined clutches of the two females) for each egg separately. As expected, in these clutches each egg seemed to be brooded in only $78.2 \pm 0.04\%$ ($n = 37$ eggs in 6 nests) of our nest visits, and most of the eggs were brooded irregularly. Incomplete brooding was rarely observed in monogamous nests (0.03% of nest visits, $n = 8462$). In these nests, reduced nest attendance accounted

for bad brooding, particularly in the small clutches (Fig. 2). Similarly, AP females without male help attended their clutch badly (Fig. 2). We tested the hypothesis experimentally by giving oystercatchers more than four eggs to brood. Again, four eggs were incubated at a time, and any surplus eggs were left cold (Fig. 2). Thus, CP trios are unable to brood more than four eggs effectively. This leads to hypothermia- and hyperthermia-induced^{14,15} retarded embryo growth and, consequently, to hatching failures. Unhatched eggs from polygynous trios showed embryo development in 47.3% of the eggs ($n = 55$), with a mean embryo size of 2.0 ± 0.3 cm ($n = 26$; egg length is approximately 5.5 cm).

Another explanation for the occurrence of polygyny could be that females expect greater benefits of settling polygynously in high-quality territories and/or with high-quality males than would be the case if they settled monogamously in low-quality territories and/or with low-quality males³. However, the number of polygynous trios was the same for low-quality territories ($0.86 \pm 0.26\%$, leapfrogs and creek leapfrogs¹⁶) and high-quality territories ($1.61 \pm 0.63\%$, residents¹⁶, $n = 14$, years 1983–1996, statistical datum derived from a Mann–Whitney U -test ($U = 94.5$, $P = 0.86$, excluding 1983–1986 when $U = 54$, $P = 0.65$). Also, monogamous CP and AP males were of equal weight and size (one-way analyses of variance (ANOVAs), d.f. = 1): weight (g) 522.1 ± 1.8 ($n = 299$), 521.0 ± 8.6 ($n = 12$), 525.4 ± 8.9 ($n = 14$, variance ratio $F = 0.08$, $P = 0.93$); wing (mm) 259.4 ± 0.4 ($n = 295$), 255.5 ± 1.6 ($n = 12$), 261.1 ± 2.0 ($n = 14$, $F = 2.3$, $P = 0.10$); tarsus (mm) 93.2 ± 0.3 ($n = 145$), 92.3 ± 0.9 ($n = 10$), 94.8 ± 0.8 ($n = 9$, $F = 1.5$, $P = 0.23$).

The fourth possible explanation is that polygyny arises because of habitat saturation³. The breeding population is very stable and there are many non-breeders competing for access to breeding territories¹⁷. Vacancies are rare⁸ and give rise to severe competition because the survival of breeders and non-breeders is very high. Females accepting a polygynous-pair bond might make the best of a bad job in most years. However, when the breeding densities are low, the percentage of polygynously paired females should decrease. In three severe winters (1986/1987, 1995/1996 and 1996/1997), the

Table 2 Reproductive performance of polygynous and monogamous males and females

Gender/Mating behaviour	Number of clutches	Laying date	Clutch size	Hatching success (%)	Fledgling production	Survival to next breeding season (%)
Males/Monogamous	1.10 ± 0.01 (1,719)	142.2 ± 0.3 (1,556)	2.50 ± 0.03 (1,719)	42.9 ± 1.1 (1,547)	0.23 ± 0.01 (1,729)	91.7 ± 0.5 (2,624)
Males/Cooperative polygynous	1.36 ± 0.15 (22)	137.0 ± 2.3 (20)	4.41 ± 0.41 (22)	21.9 ± 7.3* (20)	0.16 ± 0.16 (19)	85.0 ± 8.2 (20)
Males/Aggressive polygynous	1.00 ± 0.06 (27)	144.6 ± 2.3 (21)	2.85 ± 0.38 (27)	34.2 ± 7.8 (18)	0.04 ± 0.04 (25)	92.3 ± 5.3 (26)
Statistic <i>P</i>	3.3 0.04	2.2 0.12	29.5 <0.0001	4.7 0.09	5.9† 0.05	0.05‡ 0.97
Females/Monogamous	1.10 ± 0.01 (1,713)	142.2 ± 0.3 (1,558)	2.51 ± 0.03 (1,712)	43.0 ± 1.1 (1,547)	0.24 ± 0.01 (1,723)	92.8 ± 0.5 (2,571)
Females/Cooperative polygynous	1.16 ± 0.10 (44)	138.1 ± 1.6 (38)	2.36 ± 0.16§ (44)	23.0 ± 6.6* (34)	0.08 ± 0.06 (38)	94.9 ± 3.6 (39)
Females/Aggressive polygynous	0.75 ± 0.08 (52)	147.5 ± 2.1 (31)	1.60 ± 0.18§ (52)	25.8 ± 6.7 (30)	0.02 ± 0.02 (48)	88.2 ± 4.6 (51)
Statistic <i>P</i>	11.4 <0.0001	4.8 <0.01	16.6 <0.0001	5.7 <0.01	14.3† <0.001	1.11‡ 0.57

Sample sizes are given in parentheses. The laying date, clutch size and hatching success are depicted for the first clutch only. Significance was tested using ANOVAs with a posteriori TUKEY–HSD procedure (number of clutches, laying date and clutch size), Kruskal–Wallis with U tests (hatching success and fledgling production), and categorical logistic regression (survival). Groups not significantly different from each other are shown in bold, italics or bold italics.

* U -tests showed a significant difference between monogamous and CP nests ($Z = -2.1$, $P < 0.05$). Lower hatching success was mainly due to frequent hatching failures (HF) of eggs despite 30 days of brooding: CP nests, 51.4% HF ($n = 107$ eggs); monogamous nests, 17.4% HF ($n = 4,576$ eggs); likelihood ratio = 61.8, $P < 0.00001$ (first and replacement clutches combined, 1983–1996). † Statistics shown on the adjusted values for the mean value of the year and territory quality. ‡ Categorical logistic regression with backward deletion of the terms based on the likelihood ratio. The terms are: year (1984–1997), type of mating (monogamy, CP or AP) and the interaction between year and type of mating (all marked breeders included within and outside the main study area; observation started in 1983). Males: year, d.f. = 13, $G = 226.5$, $P < 0.0001$; interaction, d.f. = 15, $G = 9$, $P = 0.87$. Females: year, d.f. = 13, $G = 181.0$, $P < 0.0001$; interaction, $G = 14.1$, d.f. = 15, $P = 0.52$. § Clutch sizes were significantly smaller for polygynous females when compared with the clutch sizes of these same individuals when paired monogamously in other breeding seasons: Wilcoxon $Z = -2.7$, $P = 0.007$, $n = 12$.

population experienced mass mortality of both breeding and nonbreeding oystercatchers (15–27% mortality) and the number of breeders decreased. Although there was a slight increase in the occurrence of polygyny between 1987 and 1995 following the severe winter of 1986/1987 ($\chi^2 = 10$, $P = 0.002$, $n = 9$), in support of our hypothesis, there was no indication of a major setback after the winters of 1995/1996 and 1996/1997 (percentage polygyny in the main study population per year from 1984 to 1997: 0.0, 2.5, 0.0, 0.0, 1.1, 2.2, 0.9, 1.8, 2.1, 3.4, 3.1, 3.4, 2.6, 2.0).

The most likely explanation for polygyny is that both females stay within the polygynous territory because they can use it as a stepping stone to a neighbouring territory. After mild winters, six trios were seen to dissolve and both females 1 and female 2 had a high chance of obtaining a breeding position (33% and 67% chance, respectively) compared with non-breeders (9% chance). This probability increased after severe winters when many vacancies arose (to 54% and 73% chance of obtaining a position, respectively; $n = 11$) compared with the increase in probability for nonbreeders to 13%. From the females' perspective, therefore, polygyny occurs when they fail to monopolize a male; the polygynous territory is used as a stepping stone to a new breeding position; and females do not benefit from cooperative polygyny, although potentially they might (for example, they might benefit from joint brooding, which has indeed been observed once, or by laying fewer eggs). Taking the males' perspective, they suffer a loss in fitness compared with monogamous males if the females do not cooperate (Table 2). We have no evidence suggesting that males are able to predict or influence the cooperative behaviour of the females, and this inability might explain the male oystercatcher's eagerness to evict female intruders^{5,9}. □

Methods

General. In our study population on Schiermonnikoog (refs 16, 17), 25–149 breeding pairs were colour-banded and studied intensively from 1983 to 1997, and another 200–500 breeding pairs outside the main area were studied less intensively. In total, 28 polygynous territories were discovered from 1983 to 1997 (including one case of polygynandry). Because of changes in territory owners, in total 32 males (28 marked) and 58 females (51 marked) were involved. For all 58 females plus 2 females who joined a breeding pair for the second time, it was established whether they acted cooperatively or aggressively. We labelled the resident or first-laying female 'female 1' and the female joining the pair or laying last 'female 2'. In 19 territories, we recorded how the trio came into existence, and in 17 cases how it dissolved. In the joint nests of CP trios, the individual eggs were attributed to each female on the basis of the individual's specific size, colour and the markings of the eggs. Reference values for these females were obtained from eggs laid monogamously in other years, or from eggs laid during an observation session and measured directly afterwards. Reproductive performance of the birds was quantified by regular nest searches and territory observations^{16,17}. The number of clutches (0–4: 1 clutch plus up to 3 replacement clutches), the laying date (1st January = 1), the clutch size, the hatching success and the fledgling production (0–4 fledged chicks; data from 1983 to 1996) were recorded. Irregular territory visits outside the main study area in 1983–1994 account for missing values for reproductive performance. Reproductive performance of the polygynous birds within and outside the main study area (in 1983–1996) was compared to the combined reproductive performance of monogamous birds within the main study area (1983–1996) and outside the main study area (1995–1996). In the statistical analysis, the reproductive parameters were also adjusted for annual and territorial quality differences^{16,17}, but gave essentially the same results.

Behavioural observations. Colour-marked breeding individuals in the prelaying and laying period of April–June 1995 were observed for 4–6.5 h from low-to-high or high-to-low tide. One to two polygynous trios and three to five monogamous pairs from the same quality of territory¹⁶ were observed simultaneously. In total, 4 CP trios, 5 AP trios ($n = 4$ –14 observations per trio) and 44 monogamous pairs ($n = 1$ –12 observations per pair) were followed. The time spent in each behaviour (aggression towards oystercatchers, resting, feeding, brooding, copulation, flying and aggression towards other species)

when the bird was within sight was recorded for each individual on a PSION hand-held computer using Observer 3.0 software. We also noted with which birds the individuals copulated and with which birds they cooperated in aggression. Copulations were considered complete when the male or the female in the male role pushed their tail under the female in the female role⁹. In total, 8–107 aggressions per female were recorded in CP trios, 12–96 in AP trios and 1–81 in monogamous pairs (in 2 pairs out of 44 monogamous pairs, no aggressions were seen).

Genetic analysis. DNA was isolated from blood samples taken from 118 different females during 1994–1996. Microsatellite products of seven polymorphic loci were amplified by the polymerase chain reaction and analysed on ~5% polyacrylamide gels, using a plasmid vector pBluescript sequence ladder as a size standard. Genetic relatedness (r_{xy}) between individuals was calculated¹⁸. The r_{xy} values between polygynous females ($n = 7$ pairs) were compared to a frequency distribution of r_{xy} values of 1,000 randomly generated pairs from the pool of 118 breeding females. As expected in a population of unrelated individuals, the r_{xy} values were approximately normally distributed, with a mean relatedness of 0 (refs 18, 19).

Incubation behaviour. Nests were located and checked every other day and new eggs were marked and measured. We noted whether the eggs were warm (brooded). The proportion of brooded eggs (data 1990–1995) was lower in CP nests (because of incomplete brooding) and AP nests without male help (because of total parental absence from the nest) compared with monogamous nests and AP nests with male help ($n = 957$, two-way ANOVA clutch size $F = 15.5$, d.f. = 6, $P < 0.0001$; type of nest $F = 5.1$, d.f. = 3, $P < 0.01$; interaction $F = 3.2$, d.f. = 6, $P < 0.01$). In 1995, five monogamous nests outside the main study area were temporarily removed and replaced with nests containing four to six hollow copper eggs (121.8 ± 0.03 g, $n = 6$); each egg contained a temperature-registration device connected to the shell (TINY-TALK, measurement range -5 to $+37.8^\circ\text{C}$, best resolution 0.16°C). Every nest received four, five, or six copper eggs for two days each, in randomized succession; thus, the total session lasted six days per nest. Every 1.6 min a temperature was logged; a maximum of 1,800 readings was obtained per two days. In total, 4 of 45 eggs (in clutch sizes 4, 4, 5 and 6) failed to log data. If the temperature was above 33°C , the egg was considered to be brooded. A Kruskal Wallis test showed equal nest attendance (one or more eggs brooded) of all five pairs and clutches (4 clutch: $94.5 \pm 3.2\%$, $n = 5$; 5 clutch: $90.5 \pm 4.5\%$, $n = 5$; 6 clutch: 96.1 ± 1.1 , $n = 5$; $\chi^2 = 0.42$, $P = 0.81$), proving successful acceptance of the copper eggs by the breeders. In clutch size 5 and 6 the eggs were brooded less regularly compared with clutch size 4, typically with one or two eggs brooded quite well and one or two eggs brooded badly. Mean temperature (sample): 4: 36.2 ± 0.8 (5), 5: 33.5 ± 0.8 (5), 6: 33.9 ± 0.5 , ANOVA $F = 4.38$, $P < 0.05$; mean duration of a non-brooding bout in minutes (sample): 4: 29.0 ± 5.9 (144), 5: 49.8 ± 5.4 (318), 6: 52.8 ± 4.1 (532), three-way ANOVA effect of eggs 1 to 4: $F = 1.75$, $P = 0.16$ and clutch size: $F = 3.69$, $P = 0.025$ and pair: $F = 0.89$, $P = 0.47$; mean % of time an egg was brooded (sample): 4: 91.7 ± 0.02 (18), 5: 71.4 ± 0.03 (24), 6: 69.3 ± 0.04 (29), two-way ANOVA effect of clutch size: $F = 12.2$, $P < 0.0001$, pair: $F = 1.8$, $P = 0.15$ and interaction: $F = 1.6$, $P = 0.14$, mean % of eggs brooded (sample): 4: 91.4 ± 0.03 (5), 5: 70.9 ± 0.10 (5), 6: 68.9 ± 0.04 (5), ANOVA $F = 9.5$, $P < 0.01$.

Received 25 November; accepted 4 December 1997.

- Jönsson, P. E. & Ålerstam, T. The adaptive significance of parental role division and sexual size dimorphism in breeding shorebirds. *Biol. J. Linn. Soc.* **41**, 301–314 (1990).
- Trivers, R. L. In *Sexual Selection and the Descent of Man* (ed. Campbell, B.), 136–179 (Aldine, Chicago, 1972).
- Davies, N. B. In *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 263–294 (Blackwell, Oxford, 1991).
- Harris, M. P. Territory limiting the size of the breeding population of the oystercatcher (*Haematopus ostralegus*)—a removal experiment. *J. Anim. Ecol.* **39**, 707–713 (1970).
- Ens, B. J. *The Social Prisoner. Causes of Natural Variation in Reproductive Success of the Oystercatcher*. Thesis, Univ. Groningen (1992).
- Ens, B. J., Briggs, K. B., Safriel, U. N. & Smit, C. J. In *The Oystercatcher. From Individuals to Populations* (ed. Goss-Custard, J. D.) 186–218 (Oxford Univ. Press, 1996).
- Briggs, K. B. Repeated polygyny by oystercatchers. *Wader Study Group Bull.* **40**, 42–44 (1984).
- Ens, B. J., Safriel, U. N. & Harris, M. P. Divorce in the long-lived and monogamous oystercatcher: incompatibility or choosing the better option? *Anim. Behav.* **45**, 1199–1217 (1993).
- Heg, D., Ens, B. J., Burke, T., Jenkins, L. & Kruijt, J. P. Why does the typically monogamous oystercatcher (*Haematopus ostralegus*) engage in extra-pair copulations? *Behaviour* **126**, 247–289 (1993).
- de Waal, F. B. M. Tension regulation and nonreproductive functions of sex in captive bonobos (*Pan paniscus*). *Nat. Geogr. Res.* **3**, 318–335 (1987).

11. Jamieson, I. G. & Craig, J. L. Male-male and female-female courtship and copulation behaviour in a communally breeding bird. *Anim. Behav.* **35**, 1251-1253 (1987).
12. Hamilton, W. D. The genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 1-16 (1964).
13. Ens, B. J. Guarding your mate and losing the egg: an oystercatcher's dilemma. *Wader Study Group Bull.* **61**, 69-70 (1991).
14. Drent, R. in *Avian Biology*, Vol. 5 (eds Farner, D. S. & King, J. R.) 333-420 (Academic, New York, 1975).
15. Webb, D. R. Thermal tolerance of avian embryos: a review. *Condor* **89**, 874-898 (1987).
16. Ens, B. J., Kersten, M., Brenninkmeijer, A. & Hulscher, J. B. Territory quality, parental effort and reproductive success of oystercatchers (*Haematopus ostralegus*). *J. Anim. Ecol.* **61**, 703-715 (1992).
17. Ens, B. J., Weissing, F. J. & Drent, R. H. The despotic distribution and deferred maturity: two sides of the same coin. *Am. Nat.* **146**, 625-650 (1995).
18. Queller, D. C. & Goodnight, K. F. Estimating relatedness using genetic markers. *Evolution* **43**, 258-275 (1989).
19. Blouin, M. S., Parsons, M., Lacaille, V. & Lotz, S. Use of microsatellite loci to classify individuals by relatedness. *Mol. Ecol.* **5**, 393-401 (1996).

Acknowledgements. D.H. was supported by a grant and R.v.T. by a fellowship from the Netherlands Organization for Scientific Research. We thank B. J. Ens, L. van de Zande, M. Kersten, J. B. Hulscher, A. Brenninkmeijer, K. van Oers, I. de Groot, A. Helmhout, L. Klösters, R. Krijnen, F.-J. Voogd and N. J. Dingemans for participating in the project; J. Koenes, J. Nijboer, W. Beukema and D. Visser for technical assistance; Natuurmonumenten for permission to work in their reserve; and R. H. Drent, B. J. Ens, J. B. Hulscher, J. Tinbergen and J. P. Kruijt for suggestions. The Animal Experiments Committee RuG gave permission for blood sampling and temporary clutch manipulations.

Correspondence and requests for materials should be addressed to D.H. (e-mail: d.h.heg@biol.rug.nl).

Evidence for evolutionary conservation of sex-determining genes

Christopher S. Raymond*, Caroline E. Shamu†‡, Michael M. Shen†‡, Kelly J. Seifert§, Betsy Hirsch§, Jonathan Hodgkin† & David Zarkower*

* Institute of Human Genetics and Department of Biochemistry and § Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA

† MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Most metazoans occur as two sexes. Surprisingly, molecular analyses have hitherto indicated that sex-determining mechanisms differ completely between phyla. Here we present evidence to the contrary. We have isolated the male sexual regulatory gene *mab-3* (ref. 1) from the nematode *Caenorhabditis elegans* and found that it is related to the *Drosophila melanogaster* sexual regulatory gene *doublesex* (*dsx*)². Both genes encode proteins with a DNA-binding motif³ that we have named the 'DM domain'. Both genes control sex-specific neuroblast differentiation and yolk protein gene transcription; *dsx* controls other sexually dimorphic features as well. The form of *DSX* that is found in males can direct male-specific neuroblast differentiation in *C. elegans*. This structural and functional similarity between phyla suggests a common evolutionary origin of at least some aspects of sexual regulation. We have identified a human gene, *DMT1*, that encodes a protein with a DM domain and find that *DMT1* is expressed only in testis. *DMT1* maps to the distal short arm of chromosome 9, a location implicated in human XY sex reversal⁴. Proteins with DM domains may therefore also regulate sexual development in mammals.

Sexual development in *C. elegans* is controlled by a series of global regulatory genes⁶. The terminal global regulator, *tra-1*, acts as a genetic switch that can direct all aspects of somatic sexual development^{7,8} and which is also important in germline sexual development^{8,9}. Thus, *tra-1* must control, directly or indirectly, most sex-specific genes. The *tra-1* gene probably acts through several downstream regulators, each of which controls more

restricted aspects of sexual development. An analogous situation is found in *Drosophila*, where a different pathway of global regulators sets the activity of downstream genes that include *dsx*², *fruitless*^{10,11} and *dissatisfaction*¹². Despite the widespread occurrence of sexual dimorphism across phyla, until now no sexual regulatory genes have been found to be related between distant species.

The *mab-3* gene acts downstream of *tra-1* to promote male-specific development of two tissues¹. In the peripheral nervous system, *mab-3* activity directs differentiation of sensory ray neuroblasts into peripheral sense organs (V rays). In the intestine, *mab-3* plays an entirely different role, causing repression of vitellogenin (yolk protein) gene transcription.

We mapped the location of the *mab-3* gene by determining the

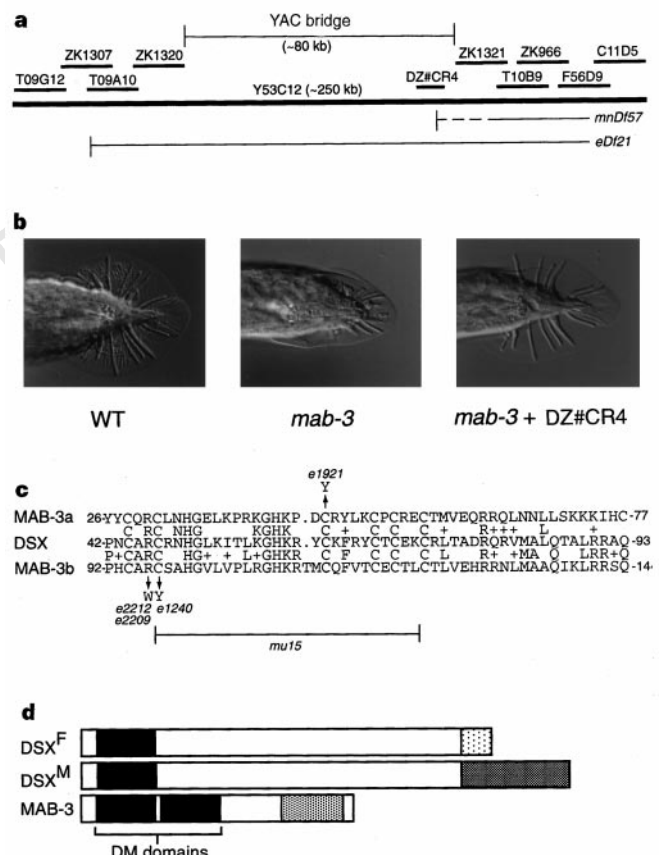


Figure 1 Isolation and characterization of the *mab-3* gene. **a**, Diagram of *mab-3* genomic region. Horizontal lines indicate positions of cosmids, phage clone DZ#CR4, and the yeast artificial chromosome clone Y53C12. *mab-3* lies between the endpoints of *eDf21* and *mnDf57*. **b**, Rescue of *mab-3* mutation by phage clone DZ#CR4. Ventral views of the tail are shown. A wild-type male tail is at the left; it has six V rays and three more posterior T rays on each side. The effect of mutant allele *e1240* is shown in the centre; T rays are present, but V rays are missing. Mutant transformed with phage clone (right) has 10 of 12 V rays restored. Nomarski optics; the original magnification was $\times 400$. **c**, Comparison of DSX and MAB-3 DM domains. Identical residues are shown in rows between protein sequences; similar residues indicated by plus signs. Amino-acid changes in *mab-3* mutants are indicated. MAB-3a is the first DM domain; MAB-3b is the second DM domain. The *mu15* allele lacks the indicated protein-coding region and most of the next intron. **d**, Diagrams of DSX and MAB-3 proteins. *dsx* is spliced to form mRNAs encoding sex-specific proteins² DSX^M and DSX^F. DM domains are indicated by filled boxes; sex-specific DSX domains and the C-terminal proline- and serine-rich region of MAB-3 are indicated by stippled boxes. The region of similarity in the DM domain corresponds closely to the minimal DNA-binding domain defined for DSX³.

‡ Present addresses: Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA (C.E.S.); Center for Advanced Biotechnology and Medicine, and Department of Pediatrics, UMDNJ-Robert Wood Johnson Medical School, 679 Hoes Lane, Piscataway, New Jersey 08854, USA (M.M.S.).

physical endpoints of genetic deficiencies (Fig. 1a). We isolated a phage clone, DZ#CR4, containing this gene. When microinjected into *mab-3* mutants, DNA from this phage clone rescues the *mab-3* mutant phenotype (Fig. 1b). We further narrowed the rescuing activity to a 7-kb interval by microinjecting DNA fragments derived from DZ#CR4 into *mab-3* mutants.

The minimal *mab-3* rescuing fragment encodes a protein with two copies of a motif similar to the DNA-binding domain of the *Drosophila* sexual regulator DSX (Fig. 1c). This domain chelates zinc and confers sequence-specific DNA binding, but is distinct from a classical zinc-finger^{3,13}. On the basis of its occurrence in DSX and MAB-3, we have named this motif the DM domain.

We sequenced DNA from the six existing *mab-3* mutant alleles and found mutations in all of them, confirming that this gene is *mab-3*. One allele, *e1921*, encodes a protein with a cysteine-to-tyrosine mutation of a conserved residue in the first DM domain (Fig. 1c), demonstrating that this domain is important for *mab-3* function. Four alleles are altered in the region encoding the second DM domain, demonstrating that this domain is also required. The canonical null allele *e1240* (refs 1, 14) encodes a protein with a cysteine-to-tyrosine mutation of a conserved residue in the second DM domain. In DSX, this cysteine is essential for DNA binding *in vitro* and for function *in vivo*³. A second null allele, *mu15*, contains a deletion that removes most of the second DM domain and the next splice junction. Two less severe loss-of-function alleles, *e2209* and *e2212*, encode proteins with the same arginine-to-tryptophan mutation. The sixth, and least severe, mutant, *e2093*, has a UAA (ochre) nonsense mutation near the 3' end, truncating the MAB-3 protein by 52 amino acids. As in DSX, the DM domains are near the amino terminus of MAB-3 (Fig. 1d).

Three mRNAs, which are formed by alternative splicing (Fig. 2a), are expressed from the *mab-3* locus. Reverse transcription with polymerase chain reaction (RT-PCR) experiments detected two *trans*-spliced messenger RNAs. The longer mRNA (~1.8 kb) encodes full-length MAB-3 protein with two DM domains. The shorter mRNA (~1.6 kb) starts just before the second DM domain. As the first AUG in this mRNA is found after the second DM domain, any protein it encodes would lack both domains. The third and shortest mRNA (~1.5 kb), identified as a cDNA clone (yk69b1; Y. Kohara, personal communication), incorporates an alternative exon with translational stops in all three frames. It may be non-functional, or it might encode a small peptide consisting of the carboxy terminus of MAB-3.

The *mab-3* mRNA is sex-specifically expressed (Fig. 2b). It is expressed in late larvae of both sexes, when *mab-3* mutations first affect the V neuroblasts¹. In the fourth larval stage, expression is approximately sixfold higher in males than in hermaphrodites. It is not clear whether the *mab-3* mRNA has a function in hermaphrodites, as *mab-3* mutations in hermaphrodites exhibit no detectable phenotype¹. In adults, expression of *mab-3* mRNA is undetectable in XX hermaphrodites, but persists in XO males. Activity of *tra-1* is required in XX animals to repress *mab-3* activity¹. Likewise, we find that *tra-1* activity is required to prevent accumulation of *mab-3* mRNA in XX animals (Fig. 2b). In adult *tra-1*(null) XX phenotypic males, *mab-3* mRNA accumulates to levels found in normal XO males.

The *mab-3* and *dsx* genes are similar not only in sequence but also in their roles in sexual development. Both genes control the differentiation of sex-specific neuroblasts into peripheral sense organs and regulate transcription of yolk protein genes. (Dipteran

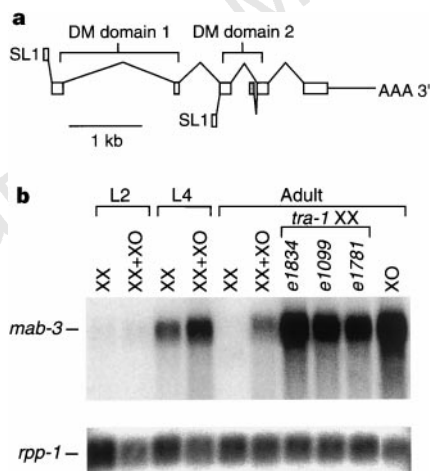


Figure 2 Splicing pattern and mRNA expression of *mab-3*. **a**, Splicing diagram. Exons are indicated by boxes and introns by lines. Spliced leader RNA SL1 is added by *trans*-splicing at positions indicated. The stippled exon contains stop codons in all three potential translational reading frames. **b**, RNA blot. Poly(A)⁺ mRNA from developmentally staged nematodes was separated by agarose gel electrophoresis and probed with a cDNA probe derived from exons 4 and 5, which are common to all three *mab-3* mRNAs. The predominant mRNAs detected are the full-length (~1.8 kb) and the shorter (~1.6 kb) *trans*-spliced mRNAs. The shortest mRNA (~1.5 kb) is not detected, is less abundant, and is not sex-specific. Chromosomal sex of animals is indicated: XX animals are wild-type hermaphrodites and XO animals are *him-8* males. XX + XO indicates mRNA from mixed sex *him-8* populations. XX *tra-1* animals are phenotypic males. The bottom panel shows a loading control in which the blot is re-probed with the *rpp-1* ribosomal protein probe. L2 and L4, larval stages.

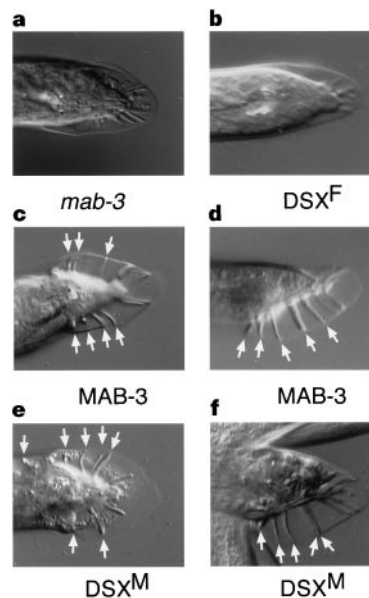


Figure 3 Male DSX, but not female DSX, can direct V-ray formation in *C. elegans*. **a**, *mab-3*(*e1240*) male with no transgene, ventral view. **b-f**, *mab-3*(*e1240*) males from transgenic strains were heatshocked. Transgenes encode DSX^F (**b**, ventral view), MAB-3 (**c**, ventral view and **d**, lateral view) or DSX^M (**e**, ventral view and **f**, ventral lateral view). V rays are indicated by white arrows. MAB-3 and DSX^M restore V rays, whereas DSX^F does not. In most animals the copulatory bursa is cupped, and it is usually not possible to photograph all rays in one focal plane. The percentages of V rays per side (wild-type animals have six rays, or 100%, per side) were as follows: *mab-3*(*e1240*) alone, 8% (240 sides counted); *e1240* expressing MAB-3 transgene, 32% (65 sides counted); *e1240* expressing DSX^M transgene, 35% (184 sides counted); *e1240* expressing DSX^F transgene, 5% (319 sides counted).

yolk proteins, however, appear to be molecularly unrelated to the vitellogenins of nematodes and vertebrates¹⁵.) Despite these similarities, *dsx* has extra functions in each sex¹⁶; thus, the two genes are not strictly equivalent. Outside the DM domain there is little sequence similarity, though the C-terminal portion of MAB-3 is proline- and serine-rich, as is the C-terminal domain of male-specific DSX protein (Fig. 1d). Rapid divergence is common in sex-determining genes. Mammalian SRY genes, for example, are highly divergent outside the HMG-box domain^{17,18}, and the human and murine SRY proteins, though they are functionally homologous, are not functionally interchangeable¹⁹. Similarly, *tra-1* and *tra-2* are the most rapidly diverging genes known in nematodes^{20,21}.

To test whether *mab-3* and *dsx* were functionally interchangeable, we examined the ability of the DSX proteins to restore V-ray formation to a *mab-3* mutant. The *dsx* gene is sex-specifically spliced to form mRNAs encoding proteins with alternative C termini (Fig. 1d)². We established transgenic *mab-3* mutant strains, containing complementary DNAs encoding MAB-3, DSX^M or DSX^F fused to heat-shock promoters. Ectopic DSX^M can restore V rays to *mab-3*(*e1240*) males essentially as well as can MAB-3 (Fig. 3). In contrast, DSX^F had no effect. These results suggest that DSX^M and MAB-3 act by similar mechanisms and can regulate similar genes in the peripheral nervous system. MAB-3 has two DM domains, whereas DSX has only one. DSX proteins dimerize¹³, however, and MAB-3 and DSX bind to similar DNA sequences *in vitro* (W. Yi and D.Z., unpublished observations).

Searches of sequence databases revealed a number of additional DM-domain-containing proteins (Fig. 4a). In *C. elegans* there are four sequenced genes, in addition to *mab-3*, that encode proteins containing a DM domain. The functions of these genes are

unknown. It is possible that they perform functions similar to those of *dsx*. Database searches also identified a human cDNA encoding a DM-domain protein, isolated from a testis cDNA library (Fig. 4a). We probed a human tissue dot blot, containing mRNA from 50 human tissues, with this cDNA. mRNA expression was detectable only in testis, which expresses an mRNA of ~2.8 kb (Fig. 4b, c). Accordingly, we have named this gene *DMT1* (for DM domain gene expressed in testis).

We mapped *DMT1* by fluorescence *in situ* hybridization (FISH) to the distal short arm of chromosome 9 (band 9p24.3; Fig. 4d). This chromosomal location has been implicated in human XY sex reversal^{4,30}. On the basis of its protein sequence, mRNA expression pattern, and map position, we propose that *DMT1* may be required in humans for male development.

Our results strongly suggest that at least some aspects of sexual regulation had a common evolutionary origin. The alternative is that the parallel roles of *dsx* and *mab-3* represent a remarkable case of convergent evolution. In contrast, the upstream global sexual regulators appear to be unrelated, at least between Nematoda and Arthropoda, where they have been most extensively characterized. Other downstream sexual regulators, for example the *fruitless*, *intersex*, *hermaphrodite* and *dissatisfaction* genes of *Drosophila*, may also be conserved across phyla. It has been proposed that the downstream regulatory genes in sex-determining pathways are the most ancient and that upstream regulatory genes were recruited subsequently²². Our data are consistent with this model. Downstream regulators such as *dsx* and *mab-3* may also be less likely to evolve because they control multiple genes: modifying such downstream regulators might also require coordinate modification of all of their targets. In either case, our results suggest that some aspects

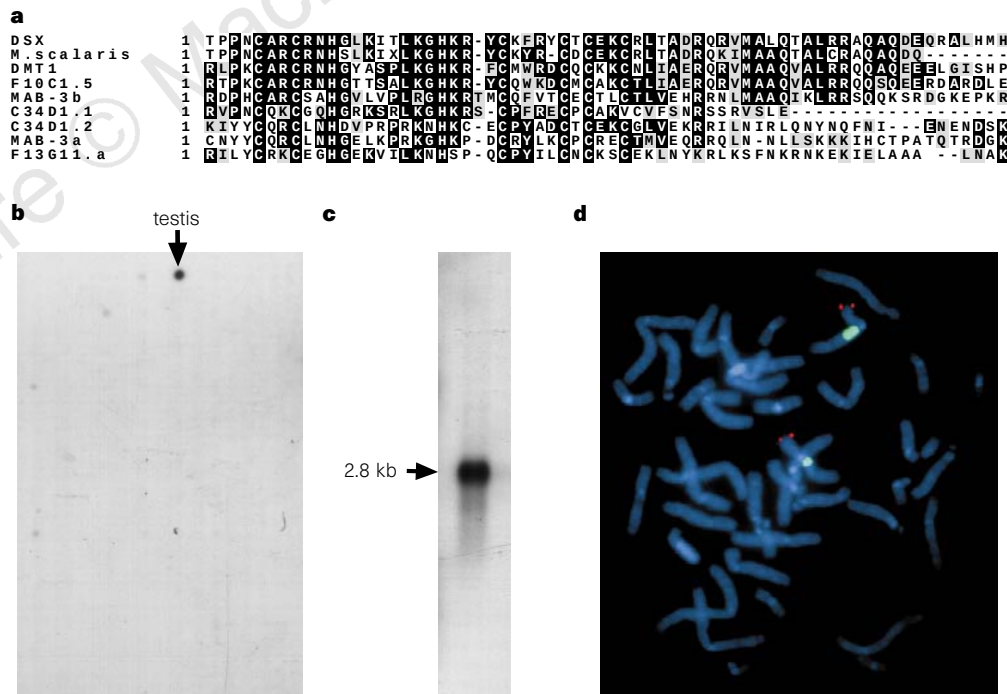


Figure 4 DM domain proteins and *DMT1*. **a**, Comparison of DM-domain-containing protein sequences. The DSX sequence is from *D. melanogaster*; the *M. scalaris* sequence is from a phorid fly (accession number Y10341); MAB-3a and MAB-3b designate the first and second DM domains of MAB-3, respectively. Other sequences are from *C. elegans* cosmids. Identical residues are shaded in black; similar residues are shaded in gray. In the *M. scalaris* sequence, X indicates an apparent frameshift in the database sequence. **b**, Testis-specific expression of *DMT1* mRNA. A human tissue dot blot with mRNA from 50 human tissues was

probed with a ³²P-labelled *DMT1* cDNA. Hybridization above background was detected only in testis mRNA (arrow). Signals in other tissues were ~10- to 80-fold lower. **c**, Northern blot of *DMT1* mRNA. Human testis poly(A)⁺ mRNA (1.5 µg) was separated by agarose gel electrophoresis and probed with a *DMT1* cDNA probe. A single mRNA of ~2.8 kb is detected (arrow). **d**, Fluorescence *in situ* hybridization (FISH) mapping of *DMT1* to chromosome 9p24.3. Hybridization of *DMT1* cDNA is in red; D9Z1 satellite DNA probe is in green; and chromosomes are counterstained in blue with DAPI.

of sexual regulation arose before the divergence of Nematoda and Arthropoda and have been conserved. □

Methods

C. elegans strains and culture. Culture and genetic manipulation of *C. elegans* were performed as previously described²³. Phenotypically wild-type males carried the *him-8(e1489)* mutation, which causes high incidence (~40%) of males due to X-chromosome nondisjunction. *mab-3* mutants were genotype *mab-3(e1240)*; *him-5(e1490)*. Transgenic nematodes were generated by standard methods²⁴, with the dominant transformation marker *rol-6(su1006)* at 100 ng μl^{-1} . XO males were isolated by filtration through nylon screens²⁵, yielding >95% pure XO males. *tra-1* mutant pseudomales were derived from strains of genotype *tra-1(lf)*; *eDp6* (*eDp6* is a free duplication containing a wild-type *tra-1* gene). Spontaneous loss of *eDp6* leads to ~40% homozygous *tra-1(lf)* animals in the population²⁵. These XX phenotypic males were isolated by filtration.

DNA clones and libraries. The *C. elegans* ribosomal protein gene *rp21* (*rpp-1*) cDNA clone pPD33.24 was a gift from A. Fire. The *DMT1* cDNA clone (Genbank accession number AA412330) was purchased from Genome Systems (St Louis). The *mab-3* cDNA clone yk69b1 (Genbank accession number D65822) was a gift from Y. Kohara. Accession numbers for *C. elegans* DM-domain-containing genes are as follows: *mab-3* (AF022388), F10C1.5 (U49831), C34D1.1 (Z78060A), C34D1.2 (Z78060B), and F13G11.a (Z83317). The *M. scalaris* gene accession number is Y10341. Multiple sequence alignments were performed using the PIMA 1.4 (<http://dot.imgen.bcm.tmc.edu:9331/multi-align/multi-align.html>) and BOXSHADE 3.21 (http://ulrec3.unil.ch/software/BOX_form.html) programs.

Cloning of *mab-3*. We mapped the location of *mab-3* by determining the endpoints of genetic deficiencies (Fig. 1a). *mnDf57* (ref. 26), which does not remove *mab-3*, ends between cosmids ZK1321 and ZK1320, while *eDf21* (ref. 1), which does remove *mab-3*, ends in cosmid ZK1307. These deficiencies thus defined an uncloned region of ~80–100 kb containing *mab-3*. We made a plasmid library of *EcoRI* fragments from the overlying yeast artificial chromosome clone Y53C12 in pBluescript II. We sequenced random clones, and identified those that were neither yeast contaminants nor derived from the previously sequenced ends of Y53C12. Fragments were positioned relative to the ends of existing contigs by polymerase chain reaction (PCR) using the Expand Long polymerase (Boehringer). This identified one DNA fragment located 4 kb to the left of the ZK1321 contig. We probed a λ phage library from yeast carrying Y53C12 with this fragment, identifying five overlapping clones from 10,000 p.f.u. plated. We microinjected one clone, DZ#CR4, into *mab-3(e1240)*; *him-5(e1490)* and found rescue of the *mab-3* phenotype both in F_1 progeny of injected animals and in four out of four transmitting lines carrying extrachromosomal arrays (Fig. 1b).

RT-PCR. Random-primed cDNA was made from RNA of mixed-sex L4 *him-8(e1489)* animals using the cDNA Cycle Kit (Invitrogen). To identify trans-spliced *mab-3* mRNAs, we performed reverse transcription (RT)-PCR with spliced leader SL1 and SL2 primers as described²⁷. Reverse *mab-3* primers were derived from exons 3, 4 and 5. PCR products were gel-purified and directly sequenced by [³³P]-labelled dideoxynucleotide (Amersham) cycle sequencing. This identified trans-splicing of SL1 to exons 1 and 3. One *mab-3* mRNA (represented by cDNA clone yk69b1) has an alternative exon with stop codons in all frames. RT-PCR and 5' RACE (rapid amplification of cDNA ends) PCR indicated that the alternative exon is the first exon in this splice variant. mRNA structures of the other two transcripts and full-length MAB-3 protein size were further confirmed by 5' RACE and by *in vitro* translation of *mab-3* cDNA.

Sequencing of *mab-3* alleles. Genomic DNA was isolated from six *mab-3* mutant strains as described²³. *mab-3* DNA was amplified by PCR, gel purified, and directly sequenced using [³³P]-labelled dideoxynucleotide (Amersham) cycle sequencing. The region encoding the second DM domain was sequenced from alleles *e1240*, *e2209*, *e2212* and *mu15*. Both DM domains were sequenced from alleles *e1921* and *e2093*. *e2093* is a double mutant, with a glutamine to asparagine mutation 3' to the UAA nonsense mutation.

RNA blotting. RNA isolation and blotting were performed as previously described²⁷, except that RNA was transferred to nylon membranes (Schleicher and Schuel) by pressure blotting. Blots were reprobed with an *rp21* cDNA probe as a loading control²⁸. The multiple human tissue blot was obtained from

Clontech (catalogue number 7770-01) and probed according to manufacturer's instructions². Tissues include ovary and uterus. Human poly(A)⁺ testis mRNA was purchased from Clontech. Tissue-blot hybridization was quantified using a Molecular Dynamics 445 SI phosphorimager and IPLab Gel software (Signal Analytics). As a control, the blot was probed for a ubiquitously expressed transcription factor, confirming that all spots contain intact mRNA.

Heatshock experiments. *mab-3*, *dsx*^F and *dsx*^F cDNAs were cloned into the heatshock vectors pPD49.78 and pPD49.83 (gifts from A. Fire) and co-injected into *mab-3(e1240)*; *him-5(e1490)* hermaphrodites with pRF4, both at 0.1 mg ml^{-1} . Results shown in Fig. 3 are from pPD49.78 constructs. Animals with extrachromosomal transgene arrays were allowed to lay eggs for 12 h at 20 °C. Progeny were raised at 20 °C for 24 h and heatshocked for 60 min at 33 °C every 12 h until they reached adulthood. V-ray formation was scored by Nomarski microscopy. Six independent lines with each cDNA were tested. Composition of arrays was confirmed by PCR.

Fluorescence *in situ* hybridization (FISH) mapping. Metaphase cells from peripheral blood lymphocytes of karyotypically normal human males were prepared by standard cytogenetic methods. FISH²⁹ was performed using a 1.5 kb *DMT1* cDNA fragment labelled by nick translation with digoxigenin-11-dUTP (Boehringer). Cells were co-hybridized with the biotin-labelled chromosome-9 probe D9Z1 (Oncor) and counterstained with 4,6-diamidino-2-phenylindole (DAPI). Probes were detected with rhodamine-labelled anti-digoxigenin antibodies and fluorescein isothiocyanate (FITC)-avidin (Oncor). 40 metaphase cells were examined using a triple-pass (DAPI/FITC/rhodamine) filter and images were captured using CytoVision software (Applied Imaging). In all cells examined, hybridization was detected on the distal short arm of both chromosomes 9. No signal was detected on any other chromosome. On the basis of the very distal position of *DMT1*, it was assigned to band 9p24.3. FISH results were confirmed in three separate experiments. Data from the Mouse Genome Database (http://mgd.niaiaffrc.go.jp/bin/query_homology) were used to compare the locations of *DMT1* and *tda-1*.

Received 17 July; accepted 27 October 1997.

- Shen, M. M. & Hodgkin, J. *mab-3*, a gene required for sex-specific yolk protein expression and a male-specific lineage in *C. elegans*. *Cell* **54**, 1019–1031 (1988).
- Burtis, K. C. & Baker, B. S. *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56**, 997–1010 (1989).
- Erdman, S. E. & Burtis, K. C. The *Drosophila* doublesex proteins share a novel zinc finger related DNA binding domain. *EMBO J.* **12**, 527–535 (1993).
- Vetia, R. *et al.* Deletions of distal 9p associated with 46, XY male to female sex reversal: definition of the breakpoints at 9p23.3–p24.1. *Genomics* **41**, 271–274 (1997).
- Eicher, E. M. *et al.* Sex-determining genes on mouse autosomes identified by linkage analysis of C57BL/6J-Y^{POS} sex reversal. *Nature Genet.* **14**, 206–209 (1996).
- Hodgkin, J. in *The Nematode Caenorhabditis elegans* (ed. Wood, W. B.) 243–279 (Cold Spring Harbor Laboratory Press, New York, 1988).
- Hodgkin, J. A. & Brenner, S. Mutations causing transformation of sexual phenotype in the nematode *Caenorhabditis elegans*. *Genetics* **86**, 275–287 (1977).
- Hodgkin, J. A genetic analysis of the sex-determining gene, *tra-1* in the nematode *Caenorhabditis elegans*. *Genes Dev.* **1**, 731–745 (1987).
- Schedl, T., Graham, P. L., Barton, M. K. & Kimble, J. Analysis of the role of *tra-1* in germline sex determination in the nematode *Caenorhabditis elegans*. *Genetics* **123**, 755–769 (1989).
- Ito, H. *et al.* Sexual orientation in *Drosophila* is altered by the *satori* mutation in the sex-determination gene *fruitless* that encodes a zinc finger protein with a BTB domain. *Proc. Natl Acad. Sci. USA* **93**, 9687–9692 (1996).
- Ryner, L. C. *et al.* Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* **87**, 1079–1089 (1996).
- Finley, K. D., Taylor, B. J., Milstein, M. & McKeown, M. *dissatisfaction*, a gene involved in sex-specific behavior and neural development of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **94**, 913–918 (1997).
- Erdman, S. E., Chen, H.-J. & Burtis, K. C. Functional and genetic characterization of the oligomerization and DNA binding properties of the *Drosophila* Doublesex proteins. *Genetics* **144**, 1639–1652 (1996).
- Hodgkin, J. Male phenotypes and mating efficiency in *Caenorhabditis elegans*. *Genetics* **103**, 43–64 (1983).
- Spieth, J., Denison, K., Kirtland, S., Cane, J. & Blumenthal, T. The *Caenorhabditis elegans* vitellogenin genes: short sequence repeats in the promoter regions and homology to the vertebrate genes. *Nucleic Acids Res.* **13**, 5283–5295 (1985).
- Baker, B. S. & Ridge, K. A. Sex and the single cell. I. On the action of major loci affecting sex determination in *Drosophila melanogaster*. *Genetics* **94**, 383–423 (1980).
- Whitfield, L. S., Lovell-Badge, R. & Goodfellow, P. N. Rapid sequence evolution of the mammalian sex-determining gene *SRY*. *Nature* **364**, 713–715 (1993).
- Tucker, P. K. & Lundrigan, B. L. Rapid evolution of the sex determining locus in Old World mice and rats. *Nature* **364**, 715–717 (1993).
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell, B. R. Male development of chromosomally female mice transgenic for *Sry*. *Nature* **351**, 117–121 (1991).
- Kuwabara, P. E. Interspecies comparison reveals evolution of control regions in the nematode sex-determining gene *tra-2*. *Genetics* **144**, 597–607 (1996).
- de Bono, M. & Hodgkin, J. Evolution of sex determination in *Caenorhabditis*: unusually high divergence of *tra-1* and its functional consequences. *Genetics* **144**, 587–595 (1996).
- Wilkins, A. S. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *BioEssays* **17**, 71–77 (1995).

23. Sulston, J. & Hodgkin, J. in *The Nematode Caenorhabditis elegans* (ed. Wood, W. W.) 587–606 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988).
24. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
25. Hodgkin, J. More sex-determination mutants of *Caenorhabditis elegans*. *Genetics* **96**, 649–664 (1980).
26. Sigurdson, D. C., Spanier, G. J. & Herman, R. K. *Caenorhabditis elegans* deficiency mapping. *Genetics* **108**, 331–345 (1984).
27. Zarkower, D. & Hodgkin, J. Molecular analysis of the *C. elegans* sex-determining gene *tra-1*: a gene encoding two zinc finger proteins. *Cell* **70**, 237–249 (1992).
28. Spieth, J., Shim, Y. H., Lea, K., Conrad, R. & Blumenthal, T. *elt-1*, an embryonically expressed *Caenorhabditis elegans* gene homologous to the GATA transcription factor family. *Mol. Cell Biol.* **11**, 4651–4659 (1991).
29. Trask, B. in *Genome Analysis: a Laboratory Manual* (eds Birren, B., Green, E., Hieter, P. & Myers, R.) (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, in the press).
30. McDonald, M. T., Flejter, W., Sheldon, S., Putzi, M. J. & Gorski, J. L. XY sex reversal and gonadal dysgenesis due to 9p24 monosomy. *Am. J. Med. Gen.* **73**, 321–326 (1997).

Acknowledgements. We thank colleagues at the University of Minnesota and MRC-LMB for discussions; M. de Bono for microinjection of cosmid from the *mab-3* region and for discussion; K. Burtis for *dsx* cDNAs; A. Coulson, J. Sulston, S. Chisoe and the *C. elegans* Genome Sequencing Consortium for assistance with physical mapping and sequencing of the *mab-3* region; C. Kenyon, C. Hunter and D. Cowing for the *mab-3(mul5)* allele; Y. Kohara for a *mab-3* cDNA clone; M. Sanders for a human-tissue blot; E. Parker for technical support; and V. Bardwell, J. Heasman, H. Towle, B. Van Ness and C. Wylie for critical reading of the manuscript. Some of the work by M.M.S. was performed in the laboratory of I. Greenwald, whom we thank. This work was supported by grants from the Minnesota Medical Foundation, University of Minnesota Graduate School and the NIH to D.Z., by NSF predoctoral fellowships to C.E.S. and M.M.S., and by the MRC and the HHMI.

Correspondence and requests for materials should be addressed to D.Z. (e-mail: zarkower@gene.med.umn.edu). Genbank accession number of the *mab-3* genomic sequence is AF022388.

Three neural tubes in mouse embryos with mutations in the T-box gene *Tbx6*

Deborah L. Chapman & Virginia E. Papaioannou

Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, 701 W. 168th Street, New York, New York 10032, USA

Somites, segmented mesodermal units of the vertebrate embryo, are the precursors of adult skeletal muscle, bone and cartilage¹. During embryogenesis, somite progenitor cells ingress through the primitive streak, move laterally to a paraxial position (alongside the body axis) and segment into epithelial somites². Little is known about how this paraxial mesoderm tissue is specified^{1,2}. We have previously described a mouse T-box gene, *Tbx6* (ref. 3), which codes for a putative DNA-binding protein^{4,5}. The embryonic pattern of expression of *Tbx6* in somite precursor cells suggests that this gene may be involved in the specification of paraxial mesoderm³. We now report the creation of a mutation in *Tbx6* that profoundly affects the differentiation of paraxial mesoderm. Irregular somites form in the neck region of mutant embryos, whereas more posterior paraxial tissue does not form somites but instead differentiates along a neural pathway, forming neural-tube-like structures that flank the axial neural tube. These paraxial tubes show dorsal/ventral patterning that is characteristic of the neural tube, and have differentiated motor neurons. These results indicate that *Tbx6* is needed for cells to choose between a mesodermal and a neuronal differentiation pathway during gastrulation; *Tbx6* is essential for the specification of posterior paraxial mesoderm, and in its absence cells destined to form posterior somites differentiate along a neuronal pathway.

We created a mutation in the mouse T-box gene *Tbx6* by deleting the initiating methionine and a portion of the T-box, which codes for the putative DNA-binding domain of the protein^{4,5}, using homologous recombination in embryonic stem (ES) cells (Fig. 1). Animals heterozygous for the mutant allele, *Tbx6*^{tm1Pa}, were viable and fertile and had no obvious abnormalities, but no homozygotes were detected among 94 offspring resulting from matings of heterozygotes. Dissection of embryos resulting from intercross matings (see Methods) revealed a class of morphologically distinct embryos, which were first evident at embryonic day of development

(e) 8.5. These embryos had elongated normally, but lacked trunk somites and had enlarged tail buds and kinked neural tubes (Fig. 2a). Vascular anomalies were observed, such as multiple haematomas in the spinal cord and tail bud and the lack of segmental arteries, although the extraembryonic vasculature, including the allantoic connection to the placenta, appeared normal. By e11.5, abnormal embryos were oedematous and one out of five was dead, as judged by the lack of a heartbeat. By e12.5, all mutant embryos were dead, presumably because of haemorrhaging of embryonic blood vessels. Genotyping of 149 embryos, aged between e9.5 and e13.5, from 19 matings of heterozygotes revealed that all the embryos with this phenotype (18%) were homozygous mutants whereas the normal embryos were heterozygous or wild type.

Histological abnormalities were first detected at e8.5 when 3 out of 25 somite-stage embryos displayed enlarged tail buds and abnormalities in somite formation. Although 5–7 cranial somites were present and were differentiating into sclerotome and dermamyotome, epithelial condensations with a continuous central lumen were observed in paraxial mesoderm in more caudal regions. At e9.5 and e10.5, nine putative mutants examined had a characteristic enlarged tail bud containing a mass of undifferentiated mesenchymal cells (Fig. 2b and d), with multiple rosettes of epithelializing tissue found lateral to the neural tube (Fig. 2e) where somites would be forming in normal embryos. These condensations had central lumina which coalesced more rostrally into a single central lumen (Fig. 2f and i), resulting in two paraxial tubes extending to the level of the forelimb bud, parallel to the axial neural tube (Fig. 2h). Although these tubes were closed along much of their length, there were irregularly spaced, ventral openings where the lumen was continuous with the coelom (Fig. 2g). Areas of vascular leakage and oedema were present (Fig. 2f).

Histologically, the paraxial tubes seemed very similar to neural epithelium, and no segmentation or somite differentiation was seen

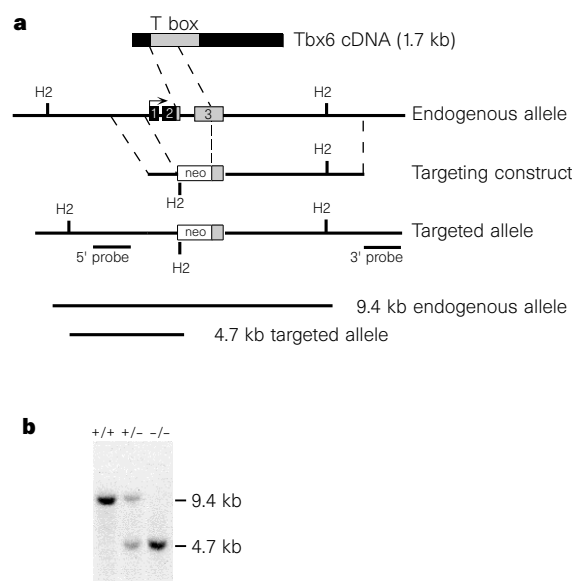


Figure 1 Targeted disruption of the *Tbx6* gene by homologous recombination. **a**, In the endogenous allele, the initiating methionine, is indicated by an arrow and the T-box is stippled. The targeting-vector construct replaces the first two exons (1, 2) and a portion of the third exon (3) with the neomycin-resistance gene (neo), in reverse transcriptional orientation. Lines below the targeted allele and probes indicate the expected sizes for *HincII* (H2)-digested endogenous and targeted alleles detected by the 5'-external probe. The 3'-external probe confirmed the targeting event. **b**, Southern-blot analysis of genomic DNA from embryos derived from matings between heterozygous mice. The 9.4-kilobase (kb) and 4.7-kb *HincII*-digested fragments corresponding to the wild-type (+) and targeted allele (-), respectively, were identified with the 5'-external probe shown in **a**.

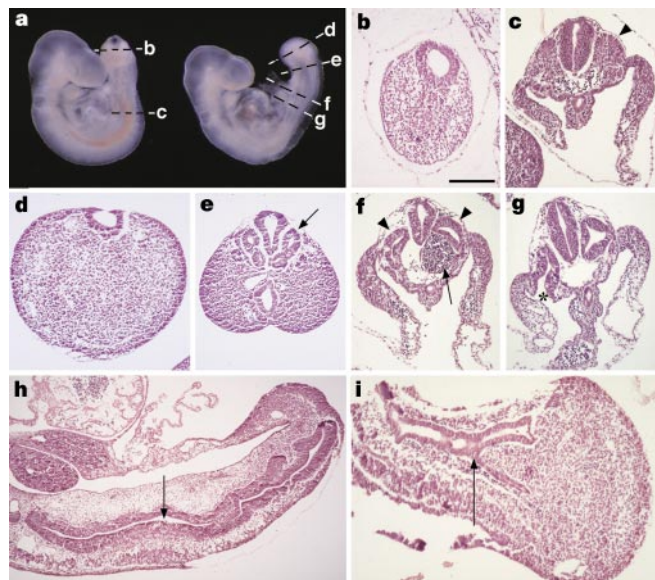


Figure 2 Normal (n) and *Tbx6*-mutant (m) embryos at e9.5 (**a–g**) and e10.5 (**h, i**). Dotted lines in **a** indicate the approximate planes of histological sections in **b–g**. **b, c**, Transverse sections of normal embryos through the tail bud and trunk, respectively. Arrowhead in **(c)** shows a differentiating somite. **d–g**, Caudal-to-rostral sections from mutant embryos. **d**, Expanded tail bud at the level of the posterior neuropore. **e**, Rosettes of epithelial cells (arrow) lateral to the closed neural tube resolve into bilateral tubes (arrowheads in **f**) with a central lumen that is continuous in some regions with the embryonic coelom (asterisk in **g**). Vascular leakage occurs as indicated by the pool of blood (arrow in **f**). **h, i**, Parasagittal sections through a mutant embryo (dorsal is at the bottom of the figures), showing a continuous paraxial tube (arrow in **h**) throughout the trunk; this tube branches in the expanded tail bud (arrow in **i**). Scale bar represents 180 μ m in **b–g**, **i**, and 600 μ m in **h**.

caudal to the forelimb bud. To characterize the molecular differentiation of the paraxial tubes, we probed for a number of markers of the primitive streak, of the mesoderm and of neural differentiation in e9.5 and e10.5 embryos. *Sonic hedgehog* (*shh*) expression⁶ confirmed the presence of a single, continuous notochord in the mutant embryos (not shown). *T*, a gene which is normally expressed in nascent mesoderm and downregulated in paraxial mesoderm⁷, was expressed throughout the expanded tail bud of mutants (Fig. 3a). *EphA4* (formerly known as *sek1*) is normally expressed in nascent mesoderm and in two transverse stripes in presomitic paraxial mesoderm, presaging the next two somites to condense^{8,9}. In mutants, *EphA4* was expressed uniformly throughout the expanded tail bud, and there was no sign of the normal striped pattern that is indicative of somite segmentation (not shown). *Wnt3a*, which is normally expressed in the primitive streak and tail bud¹⁰, is expressed in the expanded tail bud of mutants, but is more restricted to the centre than are *T* or *EphA4* (not shown).

Three other mesoderm markers, *Delta-like gene 1* (*Dll1*) (ref. 11), *paraxis*¹² and *Mox-1* (ref. 13), which are normally expressed throughout the presomitic paraxial mesoderm and upregulated in the presomitic mesoderm immediately before segmentation, were not detected in the expanded tail bud of mutants (Fig. 3b–d). Expression of *paraxis* and *Mox-1* is also characteristic of differentiating somites along the entire rostral–caudal axis in normal embryos, but expression in mutant embryos was restricted to the most anterior somites and revealed the irregular shape of these somites. There was no expression of these genes in the paraxial tubes (Fig. 3c and d). *Pax3* is a marker of presomitic paraxial mesoderm, dermamyotome, and dorsal neural tube¹⁴. In mutant embryos, the irregular anterior somites expressed *Pax3* in the dermamyotome. Caudal to the forelimb, however, there were three continuous

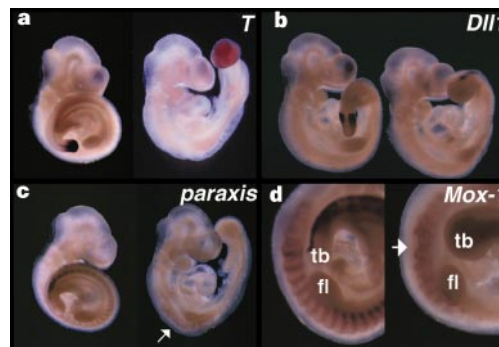


Figure 3 Expression of mesodermal marker genes in normal (left in each panel) and *Tbx6*-mutant (right in each panel) embryos at e9.5 as shown by whole-mount *in situ* hybridization (darker areas)²⁶. **a**, *T* is expressed in the normal tail bud and throughout the expanded tail of the mutant. **b**, *Delta-like gene 1* (*Dll1*), **c**, *Paraxis* and **d**, *Mox-1* are normally expressed in the presomitic mesoderm but are not expressed in this region of mutants. *Paraxis* and *Mox-1* are expressed in the irregularly shaped anterior somites (arrows), but not in the paraxial tubes, of the mutants. tb, tail bud; fl, forelimb bud.

rostrocaudal stripes of *Pax3* expression, marking the axial neural tube and the two paraxial tubes (Fig. 4a and b).

We used additional neural-tube markers to confirm that paraxial tissue in *Tbx6* mutants formed ectopic neural tubes. *HNF3 β* , which is expressed in the floorplate¹⁵, and *Pax6*, which is expressed in the lateral region of the neural tube¹⁶, were both expressed in the paraxial tubes in positions compatible with dorsal–ventral patterning relative to the notochord (ventral) and the surface ectoderm (dorsal) (Fig. 4c and d), but not in the paraxial tissue rostral to the forelimb. Differentiation within the paraxial tubes was examined using probes for *neurofilament-L* (*NF-L*)¹⁷ and *Islet1* (ref. 18). Both of these genes were expressed in basal columns in the neural and paraxial tubes, showing that motor neuron differentiation occurred in these structures (not shown). Use of *NF-L* also provided information about the segmentation of neural crest. Sensory ganglia in the head appeared normal, but disruption of segmentation of dorsal root ganglia in the cervical region and in more caudal regions was seen (Fig. 4e). It is unclear whether the paraxial tubes produce neural crest, although the histological analysis suggests that they may do so in areas where the tubes come in contact with the surface ectoderm. The expression of all of these molecular markers indicates that the paraxial tissue differentiated not as mesoderm but as neural tubes radially oriented around the notochord (Fig. 4f).

The dramatic alteration in paraxial-mesoderm differentiation seen in embryos with a mutation in *Tbx6* indicates that this putative transcription factor is needed for cells exiting the posterior streak and tail bud to become paraxial mesoderm. Cell-lineage analyses indicate that cells in the area where *Tbx6* is expressed contribute to the anterior somites and give rise to all of the posterior somite^{2,3,19}. This suggests that some cells that are destined to form anterior somites and all cells that are destined to form posterior somites transiently express *Tbx6* as they ingress through the primitive streak. We propose that *Tbx6* is involved in a signalling pathway that is essential for somite differentiation, such that anterior somites are partially dependent, and posterior somites are absolutely dependent, on transient *Tbx6* expression in the primitive streak and newly formed paraxial mesoderm. We also propose that formation of neural tubes from *Tbx6*-mutant paraxial cells indicates that neural differentiation is a default pathway for these cells and that *Tbx6* normally either inhibits a neural fate or imposes a mesodermal fate.

Two other mutations are particularly interesting in this context. A mutation in the gene encoding the signalling molecule *Wnt3a* results in a failure of paraxial mesoderm to migrate laterally out of the primitive streak; the blocked cells subsequently differentiate as a midline neural tube, located ventrally to the main neural tube²⁰.

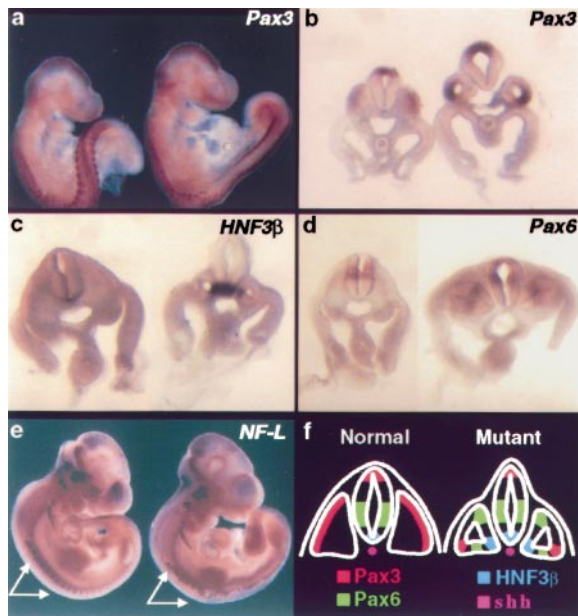


Figure 4 Expression of neural and neural/mesodermal markers in normal (left in each panel) and *Tbx6*-mutant (right in each panel) embryos at e9.5 (**a–c**) and e10.5 (**d, e**) as shown by whole-mount *in situ* hybridization (darker areas). **a**, *Pax3* is normally expressed in the dorsal neural tube and dermamyotome; in the mutant, a continuous line of expression is seen in the paraxial region and in the neural tube. **b**, A thick transverse section shows expression in the dorsal aspect of each paraxial tube. **c**, *HNF3β* and **d**, *Pax6* are normally expressed in the floor plate and lateral ventricular regions of the neural tube, respectively. In mutants, these neural markers indicate that the paraxial tubes are neural tubes radially arranged around the notochord. **e**, *Neurofilament-L* (*NF-L*) expression in the dorsal root ganglia (arrows) indicates disruption of the segmental pattern in mutants. **f**, Diagram of the expression patterns of neural-tube markers in normal and mutant embryos. *shh*, Sonic hedgehog.

Cells mutant for *fibroblast growth factor receptor-1*, when analysed in chimaeras, fail to ingress normally through the primitive streak and differentiate as ectopic neural-tube extensions of the endogenous neural tube^{21,22}. Although fibroblast growth factors (FGFs) and some T-box proteins other than *Tbx6* may function in the same signalling pathway²³, there is no evidence for direct interaction of FGF with *Tbx6* in paraxial mesoderm. The fact that *Wnt3a* is expressed in *Tbx6*-mutant embryos indicates that *Wnt3a* is not directly regulated by *Tbx6*. We propose that all three of these genes are involved in interrelated signalling pathways involved in mesoderm specification during gastrulation, such that, in the absence of any one of the gene products, mutant cells destined to become mesoderm differentiate along a neural pathway. Furthermore, the fact that the paraxial neural tubes in *Tbx6* mutants arise from cells that have undergone ingress, mesenchyme formation and lateral migration indicates that mesodermal cell fate is regulated independently of the morphogenetic movements of gastrulation. □

Methods

Mouse *Tbx6* genomic clones were isolated from a genomic λ phage 129/Sv library (Stratagene). The targeting construct was generated by replacing a 1.9-kb *SpeI/NcoI* fragment with the neomycin-resistance gene under the control of the thymidine-kinase promoter (pMC1neo polyA, Stratagene). The targeting construct was linearized at a unique *XbaI* site before insertion into electroporated R1 ES cells²⁴. We isolated DNA from cell clones and injected cells into blastocysts according to standard protocols²⁵. Chimaeric mice were mated with C57BL/6J females, and offspring carrying the mutated allele, *Tbx6*^{tm1Pa}, were crossed to produce embryos for analysis. Expression of marker genes was studied using whole-mount *in situ* hybridization²⁶.

Received 8 August; accepted 13 November 1997.

- Hogan, B., Holland, P. & Schofield, P. How is the mouse segmented. *Trends Genet.* **1**, 67–74 (1985).
- Tam, P. P. L. & Trainor, P. A. Specification and segmentation of the paraxial mesoderm. *Anat. Embryol.* **189**, 2275–2305 (1994).
- Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. & Papaioannou, V. E. *Tbx6*, a mouse T-box gene implicated in paraxial mesoderm formation at gastrulation. *Dev. Biol.* **180**, 534–542 (1996).
- Agulnik, S. I. *et al.* Evolution of mouse T-box genes by tandem duplication and cluster dispersion. *Genetics* **144**, 249–254 (1996).
- Kispert, A. & Herrmann, B. G. The *Brachyury* gene encodes a novel DNA binding protein. *EMBO J.* **12**, 3211–3220 (1993).
- Echelard, Y. *et al.* Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417–1430 (1993).
- Wilkinson, D. G., Bhatt, S. & Herrmann, B. G. Expression pattern of the mouse *T* gene and its role in mesoderm formation. *Nature* **343**, 657–659 (1990).
- Irving, C., Nieto, A. A., DasGupta, R., Charnay, P. & Wilkinson, D. Progressive spatial restriction of *Sek-1* and *Krox-20* gene expression during hindbrain segmentation. *Dev. Biol.* **173**, 26–38 (1996).
- Nieto, M. A., Gilardi-Hebenstreit, P., Charnay, P. & Wilkinson, D. G. A receptor protein tyrosine kinase implicated in the segmental patterning of the hindbrain and mesoderm. *Development* **116**, 1137–1150 (1992).
- Takada, S. *et al.* *Wnt-3a* regulates somite and tailbud formation in the mouse embryo. *Genes Dev.* **8**, 174–189 (1994).
- Bettenhausen, B., Hrabec de Angelis, M., Simon, S., Guenet, J.-L. & Gossler, A. Transient and restricted expression during mouse embryogenesis of *Dll1*, a murine gene closely related to *Drosophila Delta*. *Development* **121**, 2407–2418 (1995).
- Burgess, R., Cserjesi, P., Ligon, K. L. & Olson, E. N. Paraxis: a basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. *Dev. Biol.* **168**, 296–306 (1995).
- Candia, A. F. *et al.* *Mox-1* and *Mox-2* define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. *Development* **116**, 1123–1136 (1992).
- Williams, B. A. & Ordahl, C. P. *Pax-3* expression in segmental mesoderm marks early stages in myogenic cell specification. *Development* **120**, 785–796 (1994).
- Sasaski, H. & Hogan, B. L. M. Differential expression of multiple fork head-related genes during gastrulation and axial pattern formation in the mouse embryo. *Development* **118**, 47–59 (1993).
- Walther, C. & Gruss, P. *Pax-6*, a murine paired box gene, is expressed in the developing CNS. *Development* **113**, 1435–1450 (1991).
- Lewis, S. A. & Cowan, N. J. Genetics, evolution, and expression of the 68,000 mol. wt. neurofilament protein: isolation of a cloned cDNA probe. *J. Cell Biol.* **100**, 843–850 (1985).
- Pfaff, S. L., Mendelsohn, M., Stewart, C. L., Edlund, T. & Jessell, T. M. Requirement for LIM homeobox gene *Isl1* in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309–320 (1996).
- Tam, P. P. L. & Beddington, R. S. P. The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. *Development* **99**, 109–126 (1987).
- Yoshikawa, Y., Fujimori, T., McMahon, A. P. & Takada, S. Evidence that absence of *Wnt-3a* signaling promotes neuralization instead of paraxial mesoderm development in the mouse. *Dev. Biol.* **183**, 234–242 (1997).
- Deng, C. *et al.* Fibroblast growth factor receptor-1 (FGFR-1) is essential for normal neural tube and limb development. *Dev. Biol.* **185**, 42–54 (1997).
- Ciruna, B. G., Schwartz, L., Harpal, K., Yamaguchi, T. P. & Rossant, J. Chimeric analysis of fibroblast growth factor receptor-1 (*Fgfr1*) function: a role for FGFR1 in morphogenetic movement through the primitive streak. *Development* **124**, 2829–2841 (1997).
- Smith, J. *Brachyury* and the T-box genes. *Curr. Opin. Genet. Dev.* **7**, 474–480 (1997).
- Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).
- Joyner, A. L. *Gene Targeting* (IRL, Oxford, New York, Tokyo, 1993).
- Wilkinson, D. G. *Whole Mount In Situ Hybridization of Vertebrate Embryos* (IRL, Oxford, 1992).

Acknowledgements. We thank D. Wilkinson, T. Jessell, A. Gossler, E. Olson and B. Hogan for probes; P. Rodriguez, S. Campbell, B. Bogosian, B. Bogosian and P. Podikis for technical assistance; and T. Jessell, C. Stern, T. Bestor, L. Silver and members of our laboratory for contributions and suggestions. This work was supported by the Raymond and Beverly Sackler Foundation and by a grant from the NIH (to V.E.P.).

Correspondence and requests for materials should be addressed to V.E.P. (e-mail: vep1@columbia.edu).

Neural noise limitations on infant visual sensitivity

Ann M. Skoczenski & Anthony M. Norcia

Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco, California 94115, USA

Visual contrast sensitivity is poor in newborn human infants, but improves rapidly to approach adult levels by 8 months of age^{1–5}. During this period, infant sensitivity can be limited by physical factors affecting photon capture, such as eye size and photoreceptor density^{6,7}. Here we show that infant visual sensitivity is also limited by high levels of noise in the neural transduction process. Using a non-invasive electrophysiological measurement^{8–10} and a visual noise titration technique¹¹, we have found that intrinsic neural noise in neonates is approximately nine times higher than in adults. As intrinsic neural noise decreases during infancy, contrast sensitivity improves proportionally, suggesting that

neural noise places critical limits on contrast sensitivity throughout development. Moreover, contrast gain control¹², an inhibitory process that adjusts visual responses to changing stimulation, is in place and operating in infants as young as 6 weeks of age, in spite of high levels of neural noise and significant immaturities in contrast sensitivity. The contrast gain control that we observed in human neonates may serve as a building block for more complex forms of visual inhibition, which develop later in infancy¹³.

Figure 1 illustrates the stimuli used in this study. Spatiotemporal noise reduces the visibility of patterns, making noise-embedded gratings (Fig. 1, right column) harder to resolve than gratings without added external noise (Fig. 1, left column). We calculated the effects of internal neural noise on contrast processing by titrating external stimulus noise and contrast thresholds. Thresholds were estimated in 32 infants and 5 adults by recording visually evoked potentials (VEPs) as a function of grating contrast for various levels of external spatiotemporal noise. Figure 2a shows examples of contrast response functions for a 1 cycle per degree sinewave grating, with and without added external noise, for three representative observers (two infants and one adult). The response to the grating stimulus increased monotonically with logarithmic contrast, and we estimated thresholds for grating visibility by extrapolating this curve to zero voltage¹⁴. This process was repeated for several levels of external noise, to produce a curve for each observer relating threshold to external noise. Threshold data and curves for the three representative observers are shown in Fig. 2b.

Each data set in Fig. 2b was fitted by a model that assumes summation of internal and external noise power¹¹. According to this model, at low levels of external stimulus noise, the grating visibility threshold is determined by internal noise (the flat portion of each curve in Fig. 2b). When external noise exceeds the effects of internal noise, threshold is determined by external noise (rising portion of

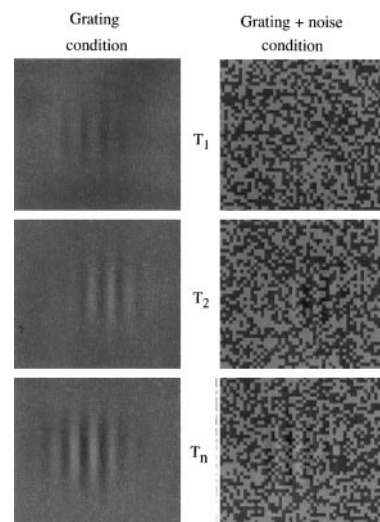


Figure 1 Schematic examples of the stimuli used in our experiments. Left column, three examples of a windowed sinewave grating with space-average luminance matched to the grey background. Grating contrast increases from top to bottom. Right column: the same three gratings with added two-dimensional noise. In our experiments, the grating stimulus was temporally modulated (contrast reversal) at a rate of 5.5 Hz to elicit visually evoked potentials. The stimulus noise was also temporally modulated, by refreshing interleaved noise frames at a rate of 33 Hz. During each 10-s 'sweep' trial, grating contrast increased logarithmically, while external noise contrast was fixed at a level between 0 and 0.40. The three left panels represent the stimulus at three points in time (T_1 , T_2 and T_n) during a trial with no external noise added, and the three right panels represent the stimulus at three points in time during a trial with fixed-contrast added external noise.

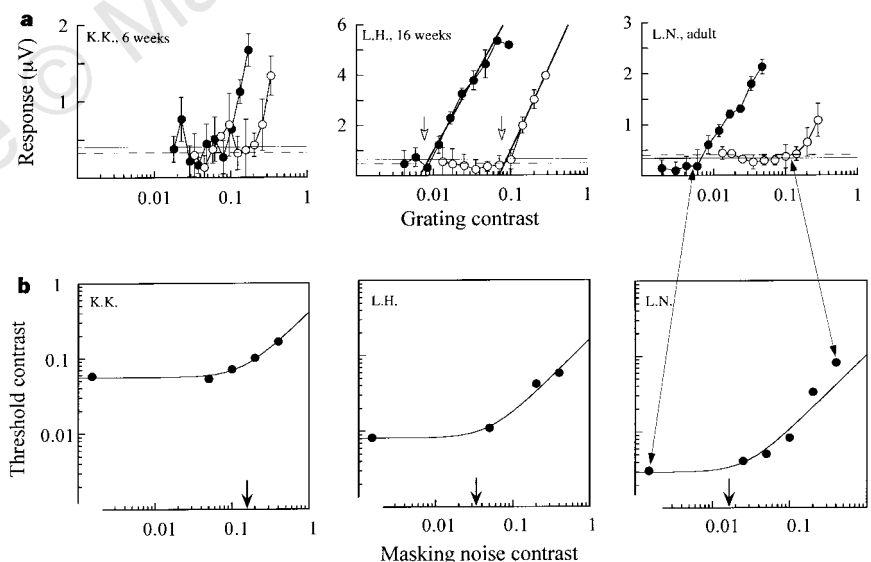


Figure 2 Data from three representative subjects: K.K., L.H. and L.N. **a**, 11.0 Hz visually evoked potential (VEP) response as a function of log grating contrast. Filled circles represent responses to the grating with no external stimulus noise added, and open circles represent responses to the grating plus external noise fixed at a contrast of 0.40. Solid and dotted horizontal lines in each plot indicate background EEG level (estimated from an average of responses at temporal frequencies just above and just below the VEP response frequency) for the no-noise and added-noise conditions, respectively. In all conditions, response increased monotonically with increasing grating contrast. Threshold was estimated by linear regression, extrapolating from the rising portion of the contrast response function to 0 µV, indicated by grey regression lines and arrows

in the plot of observer L.H.'s data (see Methods for details). **b**, Contrast thresholds as a function of external noise contrast. Each datum represents a threshold estimate derived from a contrast response function based on 6–9 10-second sweep trials (two-headed arrows between the top and bottom plots of subject L.N.'s data identify two threshold estimates, **b**, and their related contrast response functions). The left-hand point in each plot represents the subject's threshold for a grating with no added noise. The function fitted to these data points is described by the equation $T = \sqrt{N_s^2 + N_{eq}^2}$, in which T is the contrast threshold for the grating with no external noise, N_s is external noise and N_{eq} is equivalent noise (representing the effects of internal neural noise)¹¹. Each subject's N_{eq} is indicated by the arrow on the abscissa.

each curve). The intersection of these regions, called the equivalent noise (or N_{eq} , indicated by the arrow on each abscissa), indicates the level of external noise that raises the threshold by the same amount as the internal noise does. N_{eq} therefore represents the effects of internal visual noise on detection. This model provided a good fit to the data from each of our observers, and allowed us to estimate N_{eq} for individual observers.

For the youngest observer (aged 6 weeks) shown in Fig. 2b, both the contrast visibility threshold and the equivalent noise were higher than in the older infant (aged 16 weeks) and the adult. These data are representative of the developmental trends for threshold and equivalent noise in our group of observers. Figure 3a shows estimates of contrast threshold and N_{eq} as a function of age for 37 observers. The two measures follow a similar developmental timecourse, with N_{eq} consistently four to five times higher than contrast threshold throughout development.

Figure 3b shows threshold against N_{eq} for individual observers, and shows the strong relationship between the two measures. Threshold was significantly correlated with N_{eq} (correlation coefficient (r) = 0.85, n = 41, P < 0.005) across two log units of variation in contrast threshold. The best-fitting line relating the two measures had a slope of 1.04, and a non-parametric test showed that this was not significantly different from a slope of 1.0 (P > 0.05). This result is evidence that N_{eq} predicts contrast threshold, and moreover that the intrinsic neural noise represented by N_{eq} places critical limits on contrast thresholds in infants as well as adults.

What factors contribute to infants' elevated equivalent noise? The neural noise limiting our estimates of contrast threshold must arise somewhere before or within the cortical generator of the VEP. Elevated equivalent noise can, in principle, reflect reduced light capture (photon efficiency) by the infants' immature photoreceptors^{6,15}. Poor photon efficiency is unlikely, however, to contribute significantly to our measures of equivalent noise as the infant VEP contrast threshold is unaffected by retinal illuminance over a 2-log-unit range for the luminance and spatiotemporal conditions used in the present experiments¹⁶. Adult contrast thresholds are also not limited by photon capture effects under the high luminance, low spatial frequency conditions used in our experiments^{16,17}. Under these conditions, threshold is reached once a criterion signal-to-noise ratio is exceeded. Our estimate of equivalent noise reflects this internal noise level—rather than photon efficiency—as it changes during development.

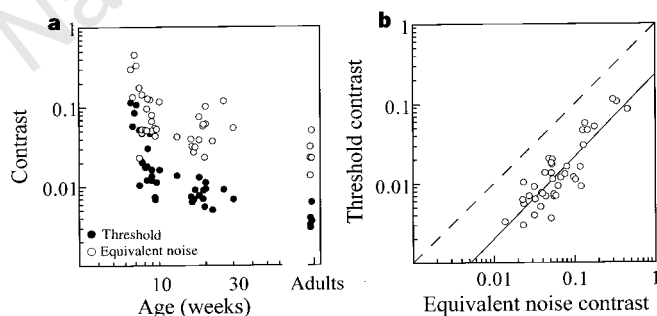


Figure 3 Relationship between contrast threshold, equivalent noise and age of subjects. **a**, Contrast threshold (filled circles) and equivalent noise contrast (open circles) as a function of age for 37 subjects. Contrast thresholds were estimated from an average of 6–9 trials. Equivalent noise was estimated from the equation in Fig. 2 fitted to 4–6 contrast thresholds measured at different levels of external noise. Both measures undergo an initial rapid phase of development between 6 and 10 weeks of age, followed by a more gradual phase of improvement. **b**, The relationship between contrast threshold and equivalent noise for individual subjects. The line of best fit to the data, with a slope of 1.1, is indicated by the solid line. The dashed line has a slope of 1.0; this slope would be expected if threshold is limited by equivalent noise.

We have also considered whether differences in background electroencephalogram (EEG) levels could have affected our measurements of equivalent noise. If EEG levels were higher in young infants, it might have been difficult to observe the visually driven signal at low contrasts. We find no significant difference, however, between background EEG levels in our youngest observers ($0.28 \pm 0.046 \mu V$) and adults ($0.22 \pm 0.11 \mu V$).

A consideration of VEP responses at suprathreshold contrasts constrains the location in the visual pathway where equivalent noise can be generated. Figure 2a shows contrast response functions with and without external noise for three observers of different ages. In all cases, high contrast external noise (at a level of 0.40) caused a rightward displacement of the contrast response function (on a log contrast axis), without altering the shape (slope) of the curve. This result implicates the functioning of a contrast scaling mechanism, a form of contrast gain control^{12,18,19}, which probably operates to maximize differential sensitivity (or visual discrimination) over the functional range of contrasts. Even the youngest infants in our sample had responses consistent with a functional contrast gain control mechanism. We see gain control effects as soon as the external noise exceeds the effects of internal neural noise, that is, for performance represented by the rising slopes of the curves in Fig. 2b. We can, therefore, infer that the neural noise limiting VEP contrast sensitivity is generated earlier in the visual pathway than the gain control mechanism.

Even with severe limits to absolute contrast sensitivity, the visual system of the neonate has an important qualitative similarity to the adult visual system: that is, the ability to adapt to changing conditions of stimulation. This early-developing contrast scaling is a form of visual inhibition: responses are reduced in the presence of high-contrast stimulation. Neural inhibition is thought to be an important factor in tuning visual responses to stimuli²⁰, and single-neuron studies of cats and monkeys have found evidence of inhibitory contrast gain control mechanisms at different stages of the visual pathway, from the retina to the cortex^{12,18,19,21,22}. The contrast gain control that we observed in human infants may serve as a building block for more complex forms of inhibition, such as orientation-specific inhibition which does not develop until 6–8 months of age¹³. □

Methods

We recorded VEPs in 37 observers: 5 adults and 32 infants ranging in age from 6 to 30 weeks of age. Twenty-eight of the infants were tested in two sessions at one age, and the remaining four infants were each tested in four sessions at two ages. Subjects had no known visual abnormalities.

Test sinusoids and spatiotemporal noise were generated on alternating video frames using look-up table animation methods. The effective frame rate for both the grating and the noise patterns was 33.12 Hz. The test gratings were orientated vertically and were reversed in contrast at 5.5 Hz using a squarewave temporal modulation profile. Grating contrast was incremented over each 10-second trial in 10 equally spaced logarithmic steps that spanned the expected contrast threshold. The noise was created by random reassignment of lookup table values at the top of each noise frame. A total of 256 noise lookup tables was generated at the beginning of each trial. These tables were selected in a random sequence during each trial with no temporal periodicity. Spatially, the noise comprised a random element pattern whose individual elements were 9.2 arcmin on a side. Display contrast was linearized over an effective 12-bit range using an attenuation network²³. Grating- and noise-frame crosstalk was measured with a fast photometer at <0.5%. Mean luminance was 100 cd m⁻² and the grating and noise patterns were co-extensive over a 15 × 20° field.

Infant fixation was monitored by an experimenter who observed the centration of the corneal reflection of the test monitor. During each 10-s trial, the experimenter interrupted data collection if the infant lost fixation. Masked and unmasked stimulus trials were interleaved.

The EEG was digitized to 16-bits accuracy at 397 Hz over a 1–100 Hz passband (–6 dB). Three occipital derivations were used: C_z versus O₁, O₂ and O₂ (international 10–20 system). The EEG obscures the observation of the

visual response as reflected in the VEP. Therefore, spectral analysis was performed using a recursive least-squares adaptive filter²⁴ to separate the VEP from the background EEG. The filter provided estimates of evoked response amplitude and phase over a 1 second running window. Six to nine individual 10-second trials were coherently averaged in the frequency domain to form averaged records of the Fourier amplitude of the second harmonic response (11 Hz) relative to the grating temporal frequency (5.5 Hz). Error statistics for averaged response amplitude were calculated using the T_{circ}^2 statistic²⁵, which also provided statistical significance values for discriminating stimulus-driven VEPs from background EEG.

We estimated contrast thresholds from the averaged records by extrapolating the response versus contrast function to 0 μV using linear regression¹⁴ (Fig. 2). The limits for the regression were selected using a combination of phase consistency and signal-to-noise ratio (SNR) criteria⁸. We calculated SNR on the basis of response amplitude and background EEG, the latter being estimated by averaging the EEG at two temporal frequencies, one just above and one just below the recording frequency. Background EEG amplitudes were not affected by visual stimulation: background levels remained constant throughout the 10-s sweep trials in both grating-alone and grating-plus-noise conditions, and the addition of external stimulus noise had no effect on background EEG levels (for example, for the youngest infants background EEG was 0.28 ± 0.046 microvolts in the grating-alone condition and 0.29 ± 0.042 microvolts in the condition with added high-contrast noise). Saturating portions of the contrast response function were excluded from the regression.

Received 15 August; accepted 4 November 1997.

- Atkinson, J., Braddick, O. & Braddick, F. Acuity and contrast sensitivity of infant vision. *Nature* **247**, 403–404 (1974).
- Harris, L., Atkinson, J. & Braddick, O. Visual contrast sensitivity of a 6-month-old infant measured by the evoked potential. *Nature* **264**, 570–571 (1976).
- Norcia, A. M., Tyler, C. W. & Hamer, R. D. High visual contrast sensitivity in the young human infant. *Invest. Ophthalmol. Vis. Sci.* **29**, 44–49 (1988).
- Norcia, A. M., Tyler, C. W. & Hamer, R. D. Development of contrast sensitivity in the human infant. *Vision Res.* **30**, 1475–1486 (1990).
- Pirchio, M., Spinelli, D., Fiorentini, A. & Maffei, L. Infant contrast sensitivity evaluated by evoked potentials. *Brain Res.* **141**, 179–184 (1978).
- Banks, M. S. & Bennett, P. J. Optical and photoreceptor immaturities limit spatial and chromatic vision of human neonates. *J. Opt. Soc. Am.* **5**, 2059–2079 (1988).
- Wilson, H. R. Development of spatiotemporal mechanisms in the human infant. *Vision Res.* **28**, 611–628 (1988).
- Norcia, A. M., Clarke, M. & Tyler, C. W. Digital filtering and robust regression techniques for estimating sensory thresholds from the evoked potential. *IEEE Engng Med. Biol.* **4**, 26–32 (1985).
- Norcia, A. M., Tyler, C. W., Hamer, R. D. & Wesemann, W. Measurement of spatial contrast sensitivity with the swept contrast VEP. *Vision Res.* **29**, 627–637 (1989).
- Tyler, C. W., Apkarian, P., Levi, D. M. & Nakayama, K. Rapid assessment of visual function: an electronic sweep technique for the pattern evoked potential. *Invest. Ophthalmol. Vis. Sci.* **18**, 703–713 (1979).
- Pelli, D. G. In *Vision: Coding and Efficiency* (ed. Blackmore, C.) 1–24 (Cambridge University Press, Cambridge, 1990).
- Bonds, A. B. Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Vis. Neurosci.* **6**, 239–255 (1991).
- Morrone, M. C. & Burr, D. C. Evidence for the existence and development of visual inhibition in humans. *Nature* **321**, 235–237 (1986).
- Campbell, F. W. & Maffei, L. Electrophysiological evidence for the existence of orientation and size detectors in the human visual system. *J. Physiol. (Lond.)* **207**, 635–652 (1970).
- Yuodelis, C. & Hendrickson, A. A qualitative and quantitative analysis of the human fovea during development. *Vision Res.* **26**, 847–855 (1986).
- Shannon, E., Skoczenski, A. M. & Banks, M. S. Retinal illuminance and contrast sensitivity in human infants. *Vision Res.* **36**, 67–76 (1996).
- Van Nes, F. L. & Bouman, M. A. Spatial modulation transfer in the human eye. *J. Opt. Soc. Am.* **57**, 401–406 (1967).
- Geisler, W. S. & Albrecht, D. G. Cortical neurons: isolation of contrast gain control. *Vision Res.* **32**, 1409–1410 (1992).
- Ohzawa, I., Sclar, G. & Freeman, R. D. Contrast gain control in the cat's visual system. *J. Neurophysiol.* **54**, 651–667 (1985).
- Burr, D. C. & Morrone, M. C. Inhibitory interactions in the human vision system revealed in pattern-evoked potentials. *J. Physiol. (Lond.)* **389**, 1–21 (1987).
- Shapley, R. M. & Victor, J. D. The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol. (Lond.)* **285**, 275–298 (1978).
- Skoczenski, A. M., O'Keefe, L. P., Kiropes, L., Tang, C., Hawken, M. J. & Movshon, J. A. Visual efficiency of macaque LGN neurons. *Neurosci. Abstr.* **20**, 7 (1994).
- Pelli, D. G. & Zhang, L. Accurate control of contrast on microcomputer displays. *Vision Res.* **31**, 1337–1350 (1991).
- Tang, Y. & Norcia, A. M. An adaptive filter for the steady-state VEP. *Electroencephalogr. Clin. Neurophysiol.* **96**, 268–277 (1994).
- Victor, J. D. & Mast, J. A new statistic for steady-state evoked potentials. *Electroencephalogr. Clin. Neurophysiol.* **78**, 378–388 (1991).

Acknowledgements. This work was supported by the National Eye Institute and the Smith-Kettlewell Eye Research Institute. We thank D. C. Burr, N. V. Graham and J. A. Movshon for discussions, and the parents of our 32 infant subjects for volunteering their time.

Correspondence and requests for materials should be addressed to A.M.S. (e-mail: sko@skivis.ski.org).

Requirement for IRF-1 in the microenvironment supporting development of natural killer cells

Kouetsu Ogasawara*, Shigeaki Hida*, Nazli Azimi†, Yutaka Tagaya†, Takeo Sato*, Taeko Yokochi-Fukuda*, Thomas A. Waldmann†, Tadatsugu Taniguchi* & Shinsuke Taki*

* Department of Immunology, Graduate School of Medicine and Faculty of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan

† Metabolism Branch, National Cancer Institute, National Institutes of Health, Building 10, 9000 Rockville Pike, Bethesda, Maryland 20892, USA

Natural killer (NK) cells are critical for both innate and adaptive immunity^{1,2}. The development of NK cells requires interactions between their progenitors and the bone-marrow microenvironment^{3–6}; however, little is known about the molecular nature of such interactions. Mice that do not express the transcription factor interferon-regulatory factor-1 (IRF-1; such mice are IRF-1^{−/−} mice) have been shown to exhibit a severe NK-cell deficiency^{7,8}. Here we demonstrate that the lack of IRF-1 affects the radiation-resistant cells that constitute the microenvironment required for NK-cell development, but not the NK-cell progenitors themselves. We also show that IRF-1^{−/−} bone-marrow cells can generate functional NK cells when cultured with the cytokine interleukin-15 (refs 9–12) and that the interleukin-15 gene is transcriptionally regulated by IRF-1. These results reveal, for the first time, a molecular mechanism by which the bone-marrow microenvironment supports NK-cell development.

In order to identify which cell population(s) is affected by IRF-1 deficiency, we first transferred bone-marrow cells from IRF-1^{−/−} mice¹³ that had been backcrossed several times with C57BL/6 (B6) mice into irradiated, H-2-compatible 129/SvJ (129) mice. In these chimaeric mice, donor-derived lymphoid cells can be distinguished from recipient lymphoid cells by the expression of Ly9.1, a marker protein that is expressed on lymphoid cells in 129 mice but not in B6 mice. As seen in Fig. 1a, irradiated 129 mice into which IRF-1^{−/−} bone-marrow cells had been transferred (IRF-1^{−/−} → 129 chimaeras) generated NK cells, which were defined by the expression of NK1.1 and lack of expression of Ly9.1 and CD3 on their surfaces. Similar numbers of NK cells were produced in IRF-1^{−/−} → 129 chimaeras and control wild-type B6 → 129 chimaeras. Mononuclear cells isolated from mice reconstituted with either IRF-1^{−/−} bone-marrow cells or control wild-type bone-marrow cells exhibited comparable cytotoxicity against NK-sensitive YAC-1 cells (Fig. 1b). Therefore, NK cells generated in IRF-1^{−/−} → 129 chimaeras are functional. Complement-dependent elimination of NK cells using a monoclonal antibody against NK1.1 molecules (which is expressed by B6 but not by 129-derived NK cells) abolished cytotoxicity against YAC-1 cells, demonstrating that donor-derived NK cells are responsible for this cytotoxicity (data not shown). Hence, these results demonstrate that bone-marrow cells in IRF-1^{−/−} mice are fully competent to differentiate into functional NK cells if an appropriate environment is provided by normal, radiation-resistant cells.

To confirm this, we transferred bone-marrow cells from 129 mice into irradiated IRF-1^{−/−} mice. In this case NK cells were detected by their expression of Ly9.1 and the interleukin-2 receptor- β (IL-2R β) chain, another marker for NK cells¹⁴. As shown in Fig. 2a, 129 → IRF-1^{−/−} chimaeras generated significantly fewer NK (Ly9.1⁺IL-2R β ⁺CD3[−]) cells than did control 129 → wild-type chi-

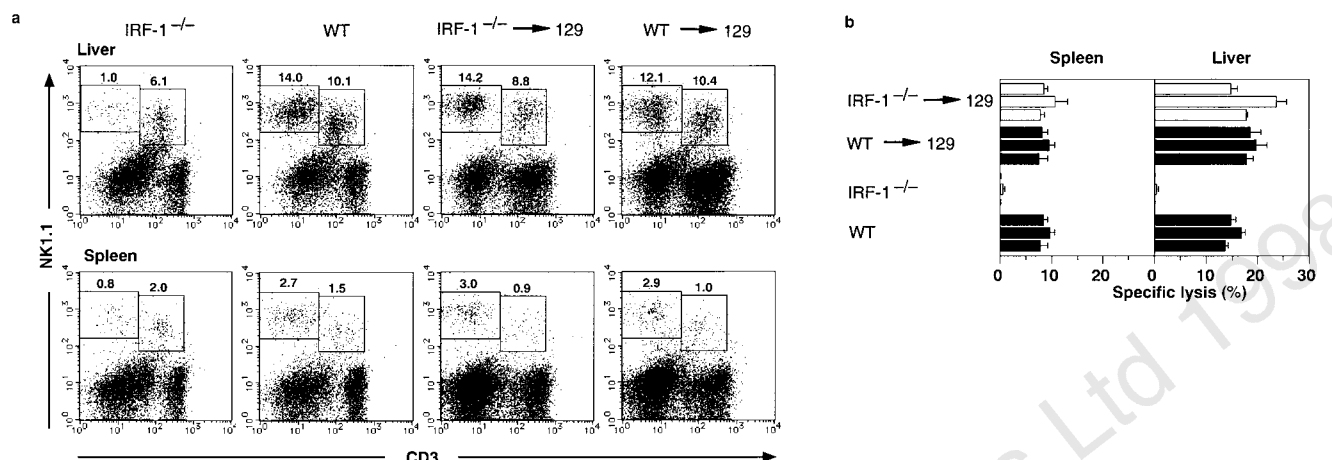


Figure 1 Generation of NK cells from bone-marrow cells upon transfer into irradiated wild-type (WT) mice. **a**, Flow cytometry analysis of the NK (NK1.1⁺CD3⁻) or NK-T (NK1.1⁺CD3⁺) cell populations in the liver (upper row) or the spleen (lower row) of IRF-1^{-/-} → 129 or WT → 129 chimaeras. Profiles representing unmanipulated IRF-1^{-/-} and wild-type mice are also included as controls. Cells expressing Ly9.1 (<10% in all cases) were excluded from the analysis. The x-axis shows the fluorescent intensity of CD3-staining and the y-axis shows the fluorescent

intensity of NK1.1-staining. The numbers in each profile indicate the percentages of cells within indicated quadrants. Representative data from one of three independent transfers are shown. **b**, Natural cytotoxicity (mean % specific lysis and s.e.m. of triplicated assays) of spleen and liver NK cells in indicated chimaeric or unmanipulated mice. Each column represents the result obtained from a single animal assayed at the effector-to-target (E/T) ratio of 50 (liver) or 200 (spleen).

maeras, and functional NK cells were totally absent in both the spleen and the liver of 129 → IRF-1^{-/-} chimaeras (Fig. 2b). We imagine that this partial restoration of the development, but not function, of NK cells may be due to the transfer of some radiation-resistant cells from the donor bone marrow. These results show that IRF-1 is dispensable for the generation of NK-cell progenitors but essential for the function of radiation-resistant cells in supporting NK-cell development. Supporting this notion, we found a similar percentage of CD45 R/B220⁺CD19⁻NK1.1⁺ cells, which may be NK-cell progenitors¹⁵, in the bone marrow of IRF-1^{-/-} and wild-type mice (data not shown).

These observations indicate the existence of an IRF-1 target gene(s) which is required to 'nurture' the development of functional NK cells. Much evidence points to the paramount role of cytokines in lymphopoiesis. In particular, interleukin (IL)-5 (refs 9–12) is seen as a key cytokine in NK-cell development, for several reasons. First, NK-cell development is impaired in mice lacking the IL-2Rβ or IL-2Rγ chains, which are shared by the IL-2 and IL-15 receptors^{16–18}. Second, NK cells develop normally in mice lacking IL-2 or the IL-2Rα chain¹⁹. Third, recombinant IL-15 (rIL-15) induces NK-cell differentiation in human bone-marrow cultures²⁰.

We therefore examined whether exogenous rIL-15 could restore the ability of IRF-1^{-/-} bone-marrow cells to create an environment that is permissive for NK-cell development. Interestingly, in the presence of rIL-15, as many functional NK cells were generated *in vitro* from IRF-1^{-/-} bone-marrow cells as were produced from wild-type bone-marrow cells (Fig. 3a and b). In contrast, IL-7, which induced B-cell differentiation from wild-type or IRF-1^{-/-} bone-marrow cells (data not shown), did not induce NK-cell differentiation (Fig. 3a and b). As IL-15 can confer cytotoxicity to nonlytic NK cells in β-oestradiol-treated mice²¹, it seems to support not only NK-cell development from bone-marrow progenitors, but also the acquisition of cytolytic activity by mature NK cells. However, NK cells generated from IL-15-supplemented IRF-1^{-/-} bone-marrow-cell culture showed significantly less lytic activity than those from wild-type bone-marrow cells (Fig. 3b). This result points to the interesting possibility that IRF-1 controls the expression of an as-yet-unidentified target gene(s) that contributes to the acquisition of cytotoxicity. Nevertheless, these results strongly suggest that the impairment of NK-cell development is due to defects in IL-15 production in IRF-1^{-/-} mice and that the *IL-15* gene is a target of IRF-1.

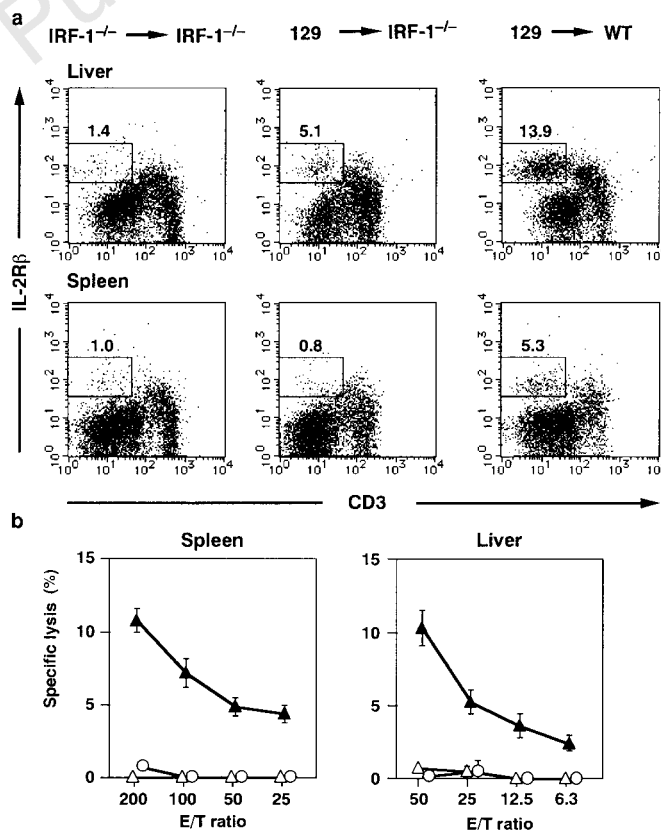


Figure 2 Inability of irradiated IRF-1^{-/-} mice to support NK-cell development from wild-type bone-marrow cells. **a**, Flow cytometry profiles for IRF-1^{-/-} mice which had received either IRF-1^{-/-} (IRF-1^{-/-} → IRF-1^{-/-}) or wild-type (129 → IRF-1^{-/-}) bone-marrow cells. The result obtained from a control transfer (129 → wild-type, WT) is also shown. Donor-derived cells constituted more than 90% of total lymphocytes. The number in each profile indicates the percentage of NK (IL-2Rβ⁺CD3⁻) cells. **b**, NK-cell-mediated cytotoxicity of IRF-1^{-/-} → IRF-1^{-/-} (open circles), 129 → IRF-1^{-/-} (open triangles) and 129 → WT (filled triangles) chimaeras. Each symbol represents the mean and the s.e.m. of triplicate assays. Representative data from one of three independent transfers are shown. We infer that the non-lytic NK cells present in 129 → IRF-1^{-/-} chimaeras developed as a result of the partial reconstitution of the environment in the host IRF-1^{-/-} mice by donor bone marrow derived cells. E/T, effector-to-target ratio.

Within the 5'-upstream region of the *IL-15* gene we found a 13-base-pair sequence, TCTTTCACCTTTC, spanning from positions -348 to -336 relative to the cap site (N.A., unpublished observations, and ref. 22). This sequence is the same as the consensus IRF-1-responsive element (IRF-E)²³. Indeed, we found that this sequence motif binds the IRF-1 protein specifically (data not shown). This IRF-E motif is critical for the activation of the *IL-15* promoter, as shown by co-transfection assays in which *IL-15*-promoter fragments, with or without this motif, were fused to the luciferase-reporter plasmid and co-transfected with an IRF-1-expression vector into P19 cells (Fig. 4a). In addition, a synthetic oligomer representing this motif alone could confer IRF-1-induced responses when ligated to a heterologous promoter in the same co-transfection assay (data not shown).

In wild-type bone-marrow cells, expression of *IL-15* messenger RNA could be strongly induced by combined stimulation with lipopolysaccharide (LPS) and interferon (IFN)- γ *in vitro*^{10,24} (Fig. 4b). In contrast, induction of transcription of the *IL-15* gene was barely detectable in IRF-1^{-/-} bone-marrow cells. Thus, IRF-1 indeed controls induction of transcription of the *IL-15* gene. Similar low-level expression of *IL-15* mRNA was observed by reverse transcription with polymerase chain reaction (PCR) analysis in both unstimulated IRF-1^{-/-} and wild-type bone-marrow cells (data not shown), indicating that IRF-1 is responsible for the induction of expression, but not the constitutive expression, of the *IL-15* gene. This observation suggests that, during the natural course of NK-cell

development, there must be some stimulus that induces *IL-15* in an IRF-1-dependent manner and which is required for development. IRF-1 is induced by a variety of stimuli, including IFN- γ (ref. 25), although this stimulus is unlikely to be IFN- γ itself as NK-cell development proceeds normally in mice lacking IFN- γ or its receptor^{26,27}. The identification of such a stimulatory ligand or factor would be an interesting avenue to follow.

It is interesting that the development of NK1.1⁺CD3⁺ cells, so-called NK-T cells²⁸, is not markedly affected in IRF-1^{-/-} mice (Fig. 1a). This finding indicates that the environment required for NK-T-cell development is distinct from that for NK-cell development with respect to the requirement for IRF-1 and *IL-15*, despite the apparently common origin of these cells^{28,29}. Identification of the signals that lead to the expression of IRF-1 and additional target genes, as well as the cell type responsible for the production of *IL-15*, will help us to dissect the intricate progenitor-environment interactions occurring during the lineage commitment of lymphoid cells. □

Methods

Bone-marrow transplantation and analysis of NK-cell development.

IRF-1^{-/-} mice used as both donors and recipients of bone-marrow transplantation had been backcrossed five times with B6 mice. 1×10^7 bone-marrow cells were transferred intravenously into irradiated mice (absorbed dose of radiation: 8.5 gray for 129 and IRF-1^{-/-} mice, and 9.0 gray for wild-type mice). Mononuclear cells isolated from the spleen and liver⁸ were analysed 8 to 10 weeks after bone-marrow transfer, by using flow cytometry with a FACScalibur and CellQuest software (Becton) with fluorochrome-conjugated monoclonal antibodies against Ly9.1, NK1.1, CD3 and IL-2R β (Pharmingen). To measure NK-mediated cytotoxicity, the standard cytotoxic assay against YAC-1 cells was performed as described previously⁸ using several different effector-to-target (E/T) ratios.

In vitro generation of NK cells. Bone-marrow cells were cultured in RPMI-

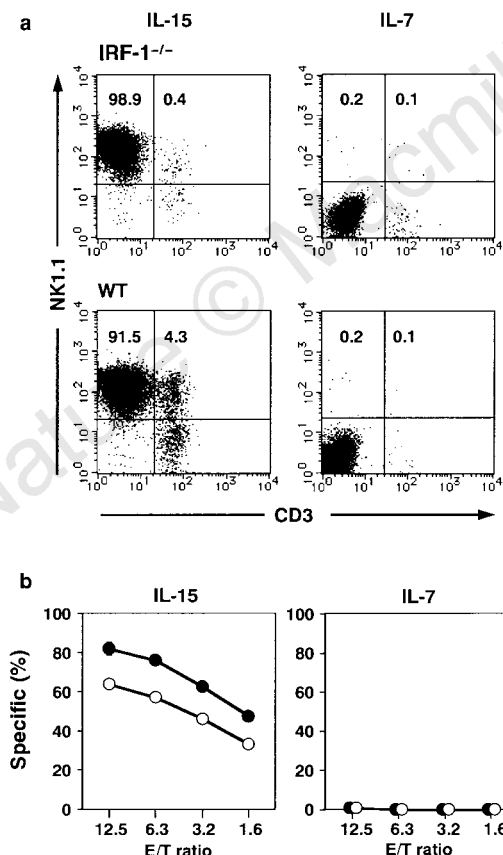


Figure 3 *In vitro* compensation of IRF-1 deficiency by *IL-15*. **a**, Generation of NK cells (NK1.1⁺CD3⁺) from IRF-1^{-/-} (upper row) or wild-type (WT; lower row) bone-marrow cells in culture supplemented with recombinant *IL-15* or recombinant *IL-7* *in vitro*. Representative data from three independent cultures are shown. **b**, Cytotoxicity of NK cells generated *in vitro* from IRF-1^{-/-} (open circles) or wild-type (filled circles) bone-marrow cells. Data represent the mean of triplicate assays. The s.e.m.s for individual determinations fall within the symbols and therefore cannot be seen. This experiment was repeated three times with similar results. E/T, effector-to-target ratio.

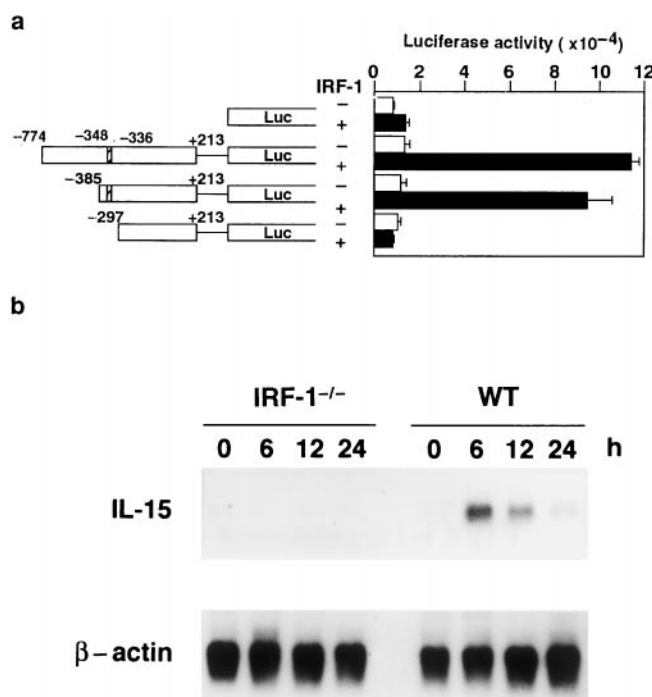


Figure 4 IRF-1 regulates the expression of the *IL-15* gene. **a**, Reporter constructs driven from the *IL-15* promoter containing or lacking the IRF-1-responsive element (cross-hatched area) were transfected together with IRF-1-expression (filled columns) or control (open columns) vectors, and luciferase (Luc) activity was measured. **b**, Expression of *IL-15* mRNA in bone-marrow cells. Total RNA prepared from IRF-1^{-/-} and wild-type (WT) bone-marrow cells which had been stimulated with IFN- γ and LPS for 0, 6, 12 and 24 h were probed for *IL-15* and β -actin mRNA expression.

1640 medium containing 10% FCS and either simian rIL-15 (Genzyme, 60 ng ml⁻¹) or murine recombinant IL-7 (R&D, 10 ng ml⁻¹) for 10 days. Cells recovered from cultures were analysed by flow cytometry or in the cytotoxicity assay.

Plasmid construction and luciferase assay. A DNA fragment spanning from positions -774 to +213 of the *IL-15* gene (N.A. and Y.T., unpublished observations) was cloned into a luciferase reporter plasmid (Promega) either directly or after deleting 5' portions. Synthetic oligomers representing the potential IRF-E in the promoter region of the *IL-15* gene were inserted along with downstream human IFN- β -promoter elements (-55 to +19 region) into the Picagene luciferase reporter plasmid (Wako)³⁰. Reporter constructs were then transfected with IRF-1-expression (pAct-1) or control (pAct-C) vectors into P19 cells, which express no endogenous IRF-1, and luciferase activity was measured as previously described³⁰.

Analysis of *IL-15* gene expression. Bone marrow cells were cultured in the presence of LPS (30 μ g ml⁻¹) and IFN- γ (100 units ml⁻¹). Total RNA was isolated by acid guanidium thiocyanate-phenol-chloroform extraction and subjected to northern blot analysis as previously described³⁰. Mouse *IL-15* and β -actin complementary DNA probes obtained by PCR were used for hybridization.

Received 16 September; accepted 24 November 1997.

- Trinchieri, G. Biology of natural killer cells. *Adv. Immunol.* **47**, 187–376 (1989).
- Scott, P. & Trinchieri, G. The role of natural killer cells in host–parasite interactions. *Curr. Opin. Immunol.* **7**, 34–40 (1995).
- Haller, O. & Wiggzell, H. Suppression of natural killer cell activity with radioactive strontium: effector cells are marrow dependent. *J. Immunol.* **118**, 1503–1506 (1977).
- Seaman, W. E., Gindhart, T. D., Greenspan, J. S., Blackman, M. A. & Talal, N. Natural killer cells, bone, and the bone marrow: studies in estrogen-treated mice and in congenitally osteopetrotic (mi/mi) mice. *J. Immunol.* **122**, 2541–2547 (1979).
- Kumar, V., Ben-Ezra, J., Bennett, M. & Sonnenfeld, G. Natural killer cells in mice treated with ⁸⁹strontium: normal target-binding cell numbers but inability to kill even after interferon administration. *J. Immunol.* **123**, 1832–1838 (1979).
- Hackett, J. Jr, Bennett, M. & Kumar, V. Origin and differentiation of natural killer cells. I. Characteristics of a transplantable NK cell precursor. *J. Immunol.* **134**, 3731–3738 (1985).
- Duncan, G. S., Mittrucker, H.-W., Kägi, D., Matsuyama, T. & Mak, T. W. The transcription factor interferon regulatory factor-1 is essential for natural killer cell function *in vivo*. *J. Exp. Med.* **184**, 2043–2048 (1996).
- Taki, S. *et al.* Multistage regulation of Th1-type immune responses by the transcription factor IRF-1. *Immunity* **6**, 673–679 (1997).
- Giri, J. G. *et al.* Utilization of the β and γ chains of the IL-2 receptor by the novel cytokine IL-15. *EMBO J.* **13**, 2822–2830 (1994).
- Grabstein, K. H. *et al.* Cloning of a T cell growth factor that interacts with the β chain of the interleukin-2 receptor. *Science* **264**, 965–968 (1994).
- Carson, W. E. *et al.* Interleukin (IL) 15 is a novel cytokine that activates human natural killer cells via components of the IL-2 receptor. *J. Exp. Med.* **180**, 1395–1403 (1994).
- Tagaya, Y., Bamford, R. N., DeFilippis, A. P. & Waldmann, T. A. IL-15: a pleiotropic cytokine with diverse receptor/signaling pathways whose expression is controlled at multiple levels. *Immunity* **4**, 329–336 (1996).
- Matsuyama, T. *et al.* Targeted disruption of IRF-1 or IRF-2 results in abnormal type I IFN gene induction and aberrant lymphocyte development. *Cell* **75**, 83–97 (1993).
- Tanaka, T. *et al.* Selective long-term elimination of natural killer cells *in vivo* by an anti-interleukin 2 receptor β chain monoclonal antibody in mice. *J. Exp. Med.* **178**, 1103–1107 (1993).
- Rolinik, A. *et al.* A subpopulation of B220⁺ cells in murine bone marrow does not express CD19 and contains natural killer cell progenitors. *J. Exp. Med.* **183**, 187–194 (1996).
- DiSanto, J. P., Muller, R., Guy-Grand, D., Fischer, A. & Rajewsky, K. Lymphoid development in mice with a targeted deletion of the interleukin 2 receptor γ chain. *Proc. Natl Acad. Sci. USA* **92**, 377–381 (1995).
- Ohbo, K. *et al.* Modulation of hematopoiesis in mice with a truncated mutant of the interleukin-2 receptor γ chain. *Blood* **87**, 956–967 (1996).
- Suzuki, H., Duncan, G. S., Takimoto, H. & Mak, T. W. Abnormal development of intestinal intraepithelial lymphocytes and peripheral natural killer cells in mice lacking the IL-2 receptor β chain. *J. Exp. Med.* **185**, 499–505 (1997).
- DiSanto, J. P. Shared receptors, distinct functions. *Curr. Biol.* **7**, R424–R426 (1997).
- Mrozek, E., Anderson, P. & Caligiuri, M. A. Role of interleukin-15 in the development of human CD56⁺ natural killer cells from CD34⁺ hematopoietic progenitor cells. *Blood* **87**, 2632–2640 (1996).
- Puzanov, I. J., Bennett, M. & Kumar, V. IL-15 can substitute for the marrow microenvironment in the differentiation of natural killer cells. *J. Immunol.* **157**, 4282–4285 (1996).
- Anderson, D. M. *et al.* Chromosomal assignment and genomic structure of IL-15. *Genomics* **25**, 701–706 (1995).
- Tanaka, N., Kawakami, T. & Taniguchi, T. Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system. *Mol. Cell. Biol.* **13**, 4531–4538 (1993).
- Bamford, R. N., Battista, A. P., Burton, J. D., Sharma, H. & Waldmann, T. A. Interleukin (IL) 15/IL-2 production by the adult T-cell leukemia cell line HuT-102 is associated with a human T-cell lymphotropic virus type I region/IL-15 fusion message that lacks many upstream AUGs that normally attenuates IL-15 mRNA translation. *Proc. Natl Acad. Sci. USA* **93**, 2897–2902 (1996).
- Taniguchi, T., Lamphier, M. S. & Tanaka, N. IRF-1: the transcription factor linking the interferon response and oncogenesis. *Biochim. Biophys. Acta* **1333**, M9–M17 (1997).
- Dalton, D. K. *et al.* Multiple defects of immune cell function in mice with disrupted interferon- γ genes. *Science* **259**, 1739–1742 (1993).
- Huang, S. *et al.* Immune response in mice that lack the interferon- γ receptor. *Science* **259**, 1742–1745 (1993).
- Bendelac, A., Rivera, M. N., Park, S. H. & Roark, J. H. Mouse CD1-specific NK1 T cells: development, specificity, and function. *Annu. Rev. Immunol.* **15**, 535–562 (1997).

- Rodewald, H. R. *et al.* A population of early fetal thymocytes expressing Fc γ RII/III contains precursors of T lymphocytes and natural killer cells. *Cell* **69**, 139–150 (1992).
- Tan, R. S., Taniguchi, T. & Harada, H. Identification of the lysyl oxidase gene as target of the antioncogenic transcription factor, IRF-1, and its possible role in tumor suppression. *Cancer Res.* **56**, 2417–2421 (1996).

Acknowledgements. We thank N. Tanaka, M. Sato and R. Perlmutter for discussion and M. S. Lamphier for critically reading the manuscript. This work was supported by the Japan Society for the Promotion of Science Research for the Future Program, by a special grant for Advanced Research on Cancer from the Ministry of Education, Science and Culture of Japan, and by the Human Frontier Science Program.

Correspondence and requests for materials should be addressed to S.T. (e-mail: shin-t@m.u-tokyo.ac.jp).

Immunoreceptor DAP12 bearing a tyrosine-based activation motif is involved in activating NK cells

Lewis L. Lanier, Brian C. Corliss, Jun Wu, Clement Leong & Joseph H. Phillips

DNAX Research Institute of Molecular and Cellular Biology, Department of Immunobiology, 901 California Avenue, Palo Alto, California 94304, USA

Natural killer (NK) cells express cell-surface receptors of the immunoglobulin and C-type lectin superfamilies that recognize major histocompatibility complex (MHC) class I peptides and inhibit NK-cell-mediated cytotoxicity¹. These inhibitory receptors possess ITIM sequences (for immunoreceptor tyrosine-based inhibitory motifs) in their cytoplasmic domains that recruit SH2-domain-containing protein tyrosine phosphatases, resulting in inactivation of NK cells^{2–4}. Certain isoforms of these NK-cell receptors lack ITIM sequences and it has been proposed that these ‘non-inhibitory’ receptors may activate, rather than inhibit, NK cells^{4–6}. Here we show that DAP12, a disulphide-bonded homodimer containing an immunoreceptor tyrosine-based activation motif (ITAM) in its cytoplasmic domain, non-covalently associates with membrane glycoproteins of the killer-cell inhibitory receptor (KIR) family without an ITIM in their cytoplasmic domain. Crosslinking of KIR–DAP12 complexes results in cellular activation, as demonstrated by tyrosine phosphorylation of cellular proteins and upregulation of early-activation antigens. Phosphorylated DAP12 peptides bind ZAP-70 and Syk protein tyrosine kinases, suggesting that the activation pathway is similar to that of the T- and B-cell antigen receptors.

It has been reported that an unknown phosphoprotein of relative molecular mass (M_r) ~12,000, expressed as a disulphide-bonded dimer, was coimmunoprecipitated from NK-cell lysates together with a non-inhibitory KIR2DS2 glycoprotein (a KIR family member with two immunoglobulin-domains in the extracellular domain, a short cytoplasmic domain lacking an ITIM, and a charged residue in the transmembrane region that is a receptor for HLA-C ligands, also referred to as p50.2 or KAR⁷). Cell-surface immunoglobulin receptors, T-cell antigen receptors (TCR), and certain Fc receptors (FcR) non-covalently associate with small transmembrane proteins (such as CD3 δ , γ , ϵ , ζ subunits, CD79 α , β , Fc ϵ RI- γ) containing ITAM sequences (D/ExxYxxL/I – x_{6–8} – YxxL/I)⁸ that are required for signal transduction by these receptor complexes⁹. Therefore, it seems likely that these non-inhibitory NK-cell receptors might require an associated protein with similar properties to mediate positive signal transduction.

A database of expressed tag sequences (EST) from a large panel of complementary DNA libraries was searched with a TBLASTN algorithm program for molecules bearing homology with the human CD3 δ , γ , ϵ , ζ and Fc ϵ RI- γ protein sequences. An EST from a human CD1+ dendritic cell library was selected for further study

because of identification of an ITAM in this molecule. Sequencing of the 604 base pair (bp) cDNA revealed an open reading frame of 339 nucleotides, encoding a putative type I membrane protein of 113 amino acids (Fig. 1a). The protein, designated DAP12, is composed of a 27 amino-acid leader, 14 amino-acid extracellular domain, 24 amino-acid transmembrane segment and 48 amino-acid cytoplasmic region. Although DAP12 has less than 25% homology with the human CD3 δ , γ , ϵ , ζ and Fc ϵ RI- γ proteins, the cytoplasmic domain contains the peptide **ESPYQELQGQRSDVYSDL** that precisely corresponds to the prototype ITAM consensus (Fig. 1b). Potential sites for phosphorylation by protein kinase C (residues 79–81 and 107–109) and casein kinase II (residues 85–88) are also present in the DAP12 cytoplasmic region. The transmembrane region contains a charged amino acid (aspartic acid (D)), also conserved in the transmembrane domain of the CD3 subunits. A potential murine homologue of DAP12 is ~70% homologous with the human DAP12 protein and has a conserved D residue in the transmembrane region, conserved cysteine (C) residues in the extracellular domain and an ITAM in the cytoplasmic region (Fig. 1c).

Southern blot analysis of human genomic DNA revealed a restriction enzyme digest pattern that is consistent with a single *DAP12* gene (Fig. 2a). *DAP12* is located on human chromosome 19q13.1. The human *KIR* genes¹⁰ and the related *LAIR*¹¹ and *ILT*/*MIR*¹² genes are all located nearby on chromosome 19q13.4. Northern blot analysis indicated the abundant presence of ~0.7 kilobase (kb) *DAP12* transcripts in human peripheral blood leukocytes and spleen, but not in thymus, prostate, testis, ovary, small intestine or colon (Fig. 2b). *DAP12* transcripts were detected in RNA isolated from two human NK cell lines NK1 and NK92, but not in the Jurkat T leukaemia cell line or the JY EBV-transformed B lymphoblastoid cell line (Fig. 2b). Southern blot analysis of a large panel of cDNA

libraries revealed predominant expression of DAP12 in resting human peripheral blood mononuclear cells, dendritic cells (from which DAP12 was cloned), peripheral blood monocytes and NK cell lines and clones (not shown).

A conspicuous feature of the non-inhibitory KIR⁵, Ly49D⁶, Ly49H⁶, CD94¹³, NKG2C¹⁴ and ILT1¹⁵ receptors is the presence of a basic amino acid (K (lysine) or R (arginine)) in the transmembrane domain. Given the precedent for interactions between proteins of multisubunit receptor complexes through oppositely charged amino acids in the transmembrane domains (such as the CD3/TCR complex)⁹, we examined whether DAP12 associates with the non-inhibitory KIR2DS2 glycoprotein containing a K in the transmembrane region¹⁶. The murine Ba/F3 pre-B cell line was transfected with a cDNA encoding KIR2DS2 either alone or together with a DAP12 cDNA containing a FLAG epitope tag at the N terminus to permit detection with an anti-FLAG monoclonal antibody. Transfectants were selected by flow cytometry for cell-surface expression based on a positive staining result with anti-KIR antibody DX27 or anti-FLAG antibody M2. KIR2DS2 Ba/F3 and KIR2DS2 + DAP12-FLAG Ba/F3 transfectants were surface-labelled with ¹²⁵I, lysed in 1% digitonin to preserve non-covalent associations of membrane protein complexes, and immunoprecipitated with anti-KIR antibody or anti-FLAG antibody. The tyrosine residue in the FLAG epitope provided a site for radioiodination, permitting visualization of the DAP12 protein. As shown in Fig. 3a, anti-KIR antibody immunoprecipitated a ¹²⁵I-labelled species of ~50–60K from both the KIR2DS2 Ba/F3 cells and KIR2DS2 + DAP12-FLAG Ba/F3 transfectants, which is consistent with the predicted molecular mass of the KIR2DS2 glycoprotein. An additional ¹²⁵I-labelled protein of ~12K was coimmunoprecipitated with anti-KIR antibody from the

a

```

50
CCACGCGTCCGCGCTGCGCCACATCCACCGGCGCTTACACTGTGGTGTG
100
CAGCAGCATCCGGCTTCATCGCGGAGTGAACCTGCGAGCGCTCCG
M C G L E P C S R L L I
150
CTCTGCGCTCTCTGCTGCGCTGTAAGTGGTCTCGTCTGTCCAGGCCA
L L P L L L A V S G L R P V Q A Q
200
GGCCAGAGCGATGCGAGTGTCTACGGTGAGCCCGGCGTGTGCGCAG
A Q S D C S C S T V S P G V L A
250
GGATCGTGTGAGGAGACCTGCTGCTGACAGTGTCTATGCGCCGCGCGT
G T V M G D L V V L T V L I A L A V
300
TACTTCTGCGCGGCTGGTCCCTCGGGGCGAGCGGCTGCGGAGGCGAG
Y P L G R L V P R G R G A A E A A
350
GACCCGGAACAGCGTATCACTGAGACCGAGTCCGCTTATCAGGAGCTCC
T R K Q R I T E T E S P Y Q E L
400
AGGTCAGAGGTCGATGTCTACAGCGACCTCAACACACAGAGCGCGTAT
Q G Q R S D V Y S D L N T Q R P Y
450
TACAAATGAGCGCAATCATGACAGTACGACACATGATACCTGGATCCAG
Y K
500
CCATCTCTGAGGCCACCCGACCTCATCTCACTCTACCGGATACATA
550
GACCCACAGAGTGCCATCCCTGAGAGACGAGACCGCTCCCAATATCTCTC
600
CTAAATAAATCATGAGCACAAAAAATAAAAAAATAAAAAAATAAAAAA
AAAA

```

b

Human DAP12-CY	-----RLVPRGRGAAEAATRKQIRITETESP-----YQELQGQRSDVYSDLNTRPYYK
Human Fc ϵ RI- γ -CY	-----RLKIQVRKAATTSYEKSDGV-----YTGLSTRNQETYEOLKHEKPP-Q
Human CD3 δ -CY	-----HETG-RLSGAADTQALLRNDQV-----YQPLRDRDDAQYSHLGGNWARNK
Human CD3 γ -CY	-----GQDGVRRQSRASDKQTLLENQDL-----YQPLKDRDDQYSHLQGNQLRRN
Human CD3 ϵ -CY	YWSKNRKAKAKPVTRCAGAGGRQGRQNKERPPVPVNPNDYEPTRKQGRDLVYSLNQRRRI
	D/E-----YxxL/I--7x-YxxL

Figure 1 DAP12 cDNA and gene. **a**, cDNA sequence of the human DAP12 cDNA (plasmid LL603, GenBank accession AF019562) and predicted polypeptide. The predicted signal peptide (amino acids 1–27) is underlined and the transmembrane (residues 42–65) is double underlined. The ITAM sequence is bold. **b**, Alignment of the cytoplasmic domains of selected human ITAM-containing polypeptides. CD3 and Fc ϵ RI- γ sequences were obtained from GenBank. **c**, Comparison of the predicted protein sequences of mouse DAP12 (GenBank accession AF024637) and human DAP12 (EC, extracellular domain; TM, transmembrane; CY, cytoplasmic domain). A murine EST (accession AA242315) with homology to human DAP12 was identified in GenBank and this cDNA was obtained and analysed by automated sequencing (ABI).

c

Human DAP12 Leader	MGGLEPCSRLLLLPLLLAVSGLRPVQA
Mouse DAP12 Leader	--A--SWC--F--V--T-G--S----
Human DAP12 EC	QAQDCSCSTVSPG
Mouse DAP12 EC	-SDFPR-D--S----
Human DAP12 TM	VLGIVMGLDLVTLVIALAVYFLG
Mouse DAP12 TM	-----L-----L-----S--
Human DAP12 CY	RLVPRGRGAAEAATRKQIRITETESPYQELQGQRSDVYSDLNTRPYYK
Mouse DAP12 CY	--S--Q-T--G-----H-A-----PE-----Q--R

KIR2DS2 + DAP12-FLAG transfectant, but not from the transfectant expressing only KIR2DS2. Reciprocally, an 125 I-labelled glycoprotein migrating in an identical way to KIR2DS2 was coimmunoprecipitated with anti-FLAG antibody from the KIR2DS2 + DAP12-FLAG Ba/F3 cells, but not from the KIR2DS2-only transfectant. Comparison of immunoprecipitates analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) using either reducing or non-reducing conditions indicate that DAP12 is expressed on the cell surface as a disulphide-bonded dimer (Fig. 3a). It should be noted that we were unable to detect cell-surface expression of DAP12 on the surface of Ba/F3 cells transfected with the DAP12-FLAG cDNA alone, without KIR2DS2. However, DAP12-FLAG proteins were detected in the cytoplasm (not shown), suggesting that DAP12 may require association with its partner subunits for efficient transport to the cell surface, in a similar manner to the situation with the CD3 proteins¹⁷. Additionally, preliminary results indicate that DAP12 does not associate with the inhibitory KIR isoforms that lack a charged residue in their transmembrane domain (unpublished observation).

A peptide corresponding to the cytoplasmic domain of DAP12 (ITETESPY*QELQGQRSDVY*SDLNTQRP) was synthesized either as an unphosphorylated protein or containing phosphates on both tyrosine (Y) residues. Lysates from Jurkat T cells or NK cell clone A6 were incubated with the biotinylated peptides and complexes precipitated using avidin-agarose. Western blot analysis demonstrated that a DAP12 peptide phosphorylated on both Y residues, but not the unphosphorylated peptide, formed complexes with the ZAP-70 kinase (Fig. 3b). The tyrosine-phosphorylated DAP12 peptide, but not the unphosphorylated DAP12 peptide, also formed a complex with the Syk protein tyrosine kinase in lysates from NK cells (not shown). The binding of these kinases to phosphorylated DAP12 is reminiscent of the interactions that have been demonstrated between the phosphorylated ITAM-containing CD3 subunits and Syk or ZAP-70 kinases during TCR signalling^{18,19}.

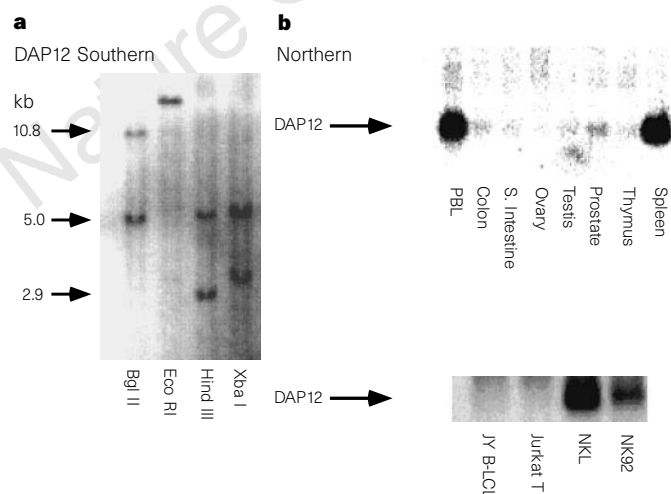


Figure 2 DAP12 gene and expression. **a**, Southern blot analysis of human genomic DNA digested with the indicated restriction enzymes and probed with the DAP12 cDNA. The restriction enzyme fragments are in accordance with the sizes predicted from the genomic sequence. The genomic organization of the human DAP12 gene (GenBank accession AF019563) was deduced from a fragment of human chromosome 19q13.1 (GenBank accession AD000833). **b**, Northern blot analysis of DAP12 in human tissues and the JY EBV-transformed B lymphoblastoid cell line, the Jurkat T leukaemia cell line, and two NK cell lines NK1 (provided by M. Robertson, Indiana University) and NK92 (obtained from H-G. Klingemann).

Ligation of the CD3/TCR complex on T cells or the immunoglobulin receptor complex on B cells results in cellular activation. Therefore, we examined the functional consequence of crosslinking the KIR2DS2-DAP12 complex. Ba/F3 transfectants expressing either KIR2DS2 alone or the KIR2DS2-DAP12-FLAG complex were incubated with anti-KIR antibody DX27 or anti-FLAG antibody, followed by a goat anti-mouse immunoglobulin to provide crosslinking. Examination of total cellular proteins in Ba/F3 cells expressing the KIR2DS2-DAP12-FLAG complex that were stimulated with anti-KIR or anti-FLAG antibodies revealed tyrosine phosphorylation of several cellular substrates (Fig. 4a). Immunoprecipitation with anti-FLAG antibody and western blot analysis with anti-phosphotyrosine antibody demonstrated that crosslinking

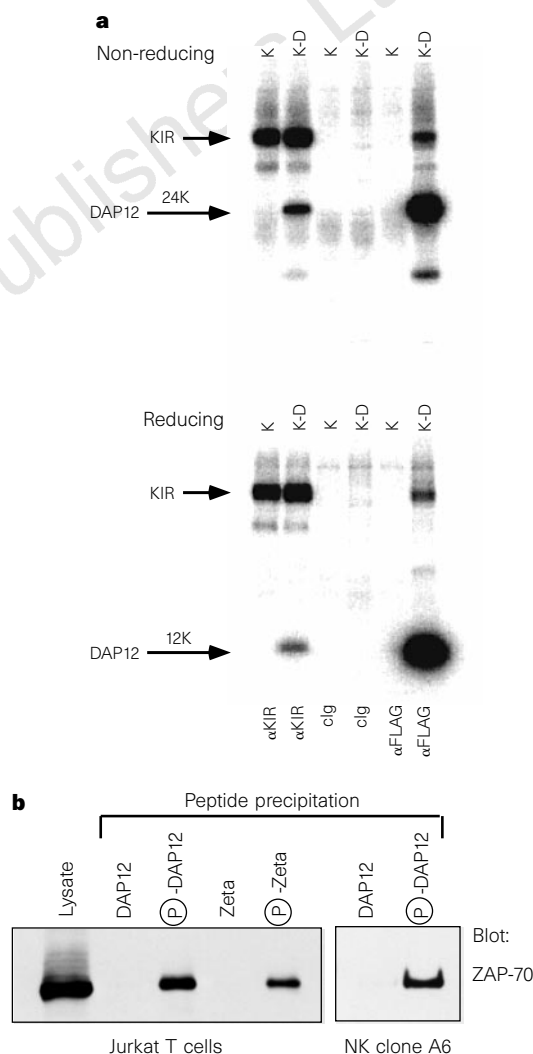


Figure 3 Coimmunoprecipitation of DAP12 and KIR2DS2 proteins and association of DAP12 phosphopeptide with ZAP-70. **a**, Murine Ba/F3 pre-B cells stably transfected with KIR2DS2 only (K) or KIR2DS2 and DAP12 (containing a FLAG epitope on the N terminus) (K-D) were labelled with 125 I, lysed in 1% digitonin buffer and antigens immunoprecipitated with control immunoglobulin (cIg), anti-KIR antibody DX27 or anti-FLAG antibody M2. Samples were analysed by SDS-PAGE using 18% acrylamide gels under non-reducing or reducing conditions. **b**, Lysates prepared from Jurkat cells or NK cell clone A6 were incubated with a biotinylated unphosphorylated (DAP12) or diphosphorylated (P-DAP12) DAP12 peptide (ITETESPY*QELQGQRSDVY*SDLNTQRP) and precipitated with avidin-agarose. Samples were analysed by western blot using anti-ZAP-70 antibody, as indicated. Biotinylated unphosphorylated or diphosphorylated CD3 ζ (P-zeta) peptides were used as a control in the experiment using lysates from Jurkat cells¹⁸.

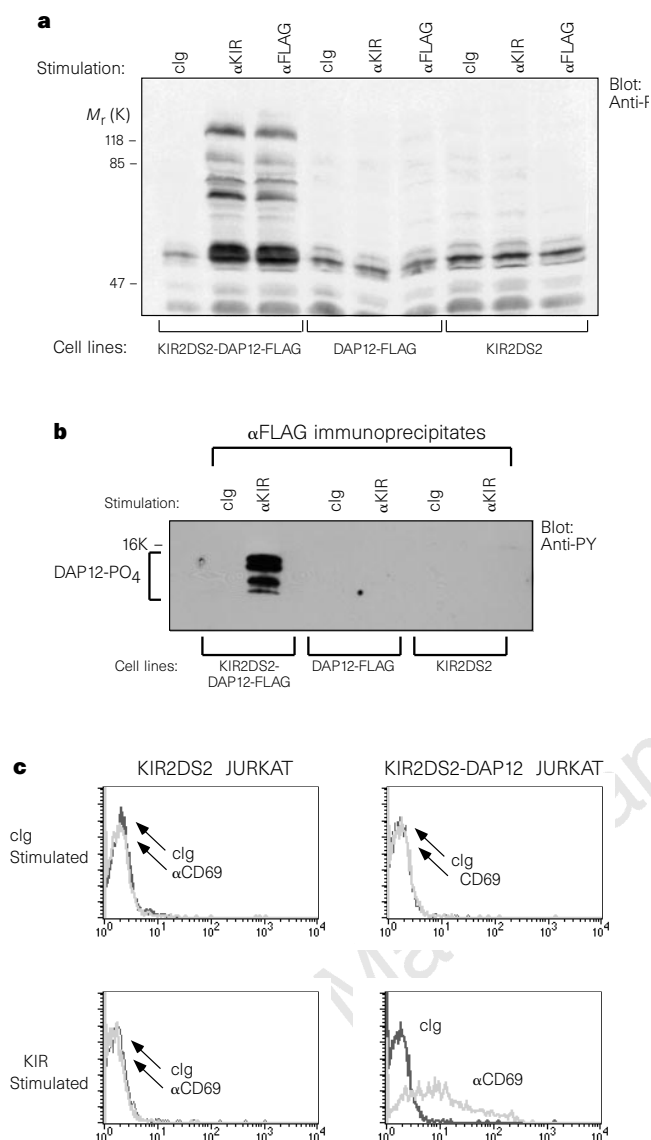


Figure 4 Cellular activation of transfectants expressing KIR2DS2 and DAP12. **a**, Total cell lysates were prepared from Ba/F3 cells transfected with KIR2DS2 alone, DAP12 (containing a FLAG epitope tag) alone or both KIR2DS2 and DAP12-FLAG that were stimulated by anti-KIR antibody DX27, anti-FLAG antibody M2 or control immunoglobulin, followed by F(ab')₂ goat anti-mouse immunoglobulin. Samples were analysed by western blot for the presence of phosphorylated cellular proteins using HRP-conjugated anti-phosphotyrosine antibody 4G10. **b**, Cell lysates were prepared from Ba/F3 cells transfected with KIR2DS2 alone or both KIR2DS2 and DAP12 (FLAG epitope tagged) that were stimulated with anti-KIR antibody DX27 or control immunoglobulin, followed by F(ab')₂ goat anti-mouse immunoglobulin. Lysates were immunoprecipitated with anti-FLAG antibody M2 or control immunoglobulin and samples were analysed by western blot for the presence of phosphorylated proteins using HRP-conjugated anti-phosphotyrosine antibody 4G10. Heterogeneity in migration of the DAP12-FLAG proteins probably reflects different phosphorylation species. **c**, Jurkat cells were stably transfected with an amphotropic retroviral vector pMX-neo²⁴ containing the NKAT5 cDNA¹⁶ encoding KIR2DS2. These KIR2DS2+ Jurkat cells were transiently transfected by electroporation with human DAP12 in the pEF-BOS vector¹⁸. After 24 h, sham (KIR2DS2 Jurkat) or DAP12 transfectants (KIR2DS2-DAP12 Jurkat) were stimulated by culture in microtitre plates precoated (5 µg ml⁻¹) with control immunoglobulin (upper panels) or anti-KIR antibody DX27 (lower panels). After 12 h incubation, cells were collected and stained with FITC-conjugated control immunoglobulin (clg) or anti-CD69 antibody, as indicated. Samples were analysed by flow cytometry (x-axis, fluorescence, 4 decade log scale; y-axis, number of cells).

the KIR2DS2–DAP12–FLAG transfectants with anti-KIR antibody induced tyrosine phosphorylation of the DAP12 protein (Fig. 4b) and resulted in the association of phosphorylated DAP12 with the Syk protein tyrosine kinase (not shown). By contrast, Ba/F3 cells expressing only KIR2DS2 were not activated by crosslinking with anti-KIR antibody. Similarly, upregulation of CD69 expression was observed in Jurkat T leukaemia cells transfected with both KIR2DS2 and DAP12, but not KIR2DS2 alone, when these receptors were crosslinked with anti-KIR antibody (Fig. 4c). These results indicate that DAP12 is necessary and responsible for KIR2DS2 signal transduction in these host cells and are in accordance with previous observations demonstrating that KIR2DS2 molecules are functional in NK cells but not in transfectants expressing only KIR2DS2²⁰.

Our studies suggest that DAP12 may associate with the non-inhibitory isoforms of the KIR molecular in NK cells and permit cellular activation through these receptors, in a similar way to the function of the CD3 subunits in the TCR complex and CD79 subunits in the B-cell-receptor complex. Whether DAP12 associates with the non-inhibitory NK-cell receptors of the murine Ly49 family and the human CD94–NKG2C and NKG2E complexes is under investigation. Expression of DAP12 in monocytes and dendritic cells predicts association with other receptors similar to the non-inhibitory KIR present in these cell types. Likely candidates are the recently identified ILT/MIR family of molecules expressed by human monocytes^{12,15} and the PIR-A molecules in rodent myeloid and B cells^{21,22}. In addition, the physical properties of DAP12 are similar to a new dimeric 12K phosphoprotein identified in the pre-T-cell-receptor complex on murine thymocytes²³. Thus, DAP12 may function in cellular activation mediated by a diverse array of receptors in distinct cell lineages.

Methods

Cloning and sequence analysis. TBLASTN searches of the DNAX sequence database were made using the human CD3δ, γ, ε, ζ and FcεRI-γ protein sequences. The cDNA insert in plasmid LL603, identified in a human CD1+ dendritic cell library, was isolated and subjected to automated sequencing (ABI).

DNA and RNA. RNA from human tissues and human genomic DNA were purchased from Clontech (Palo Alto, CA). Northern and Southern blot analysis were done as described¹³.

Transfection. A cDNA containing the CD8 leader segment, followed by the FLAG peptide epitope (DYKDDDDK), and joined to the extracellular, transmembrane and cytoplasmic segments of DAP12 was subcloned into the pMX-puro retroviral vector²⁴ (provided by T. Kitamura, DNAX), packaged using the Phoenix cell line (provided by G. Nolan, Stanford), and virus was used to infect the mouse pre-B cell line Ba/F3²⁴. The NKAT5 cDNA¹⁶ encoding KIR2DS2 (provided by M. Colonna, Basel) was subcloned into the pMX-neo retroviral vector. Ba/F3 cells were infected, drug selected, and transfectants isolated using flow cytometry²⁴. DAP12 cDNA was subcloned into the pEF-BOS vector for transient expression in Jurkat cells using electroporation for introduction of the plasmid²⁵.

Immunoprecipitation. Cells were labelled with ¹²⁵I and solubilized in lysis buffer (pH 7.8, 1% digitonin (Sigma), 0.12% Triton-X100, 150 mM NaCl, 20 mM triethanolamine, 0.01% NaN₃, and protease inhibitors)²⁶. Cell lysates were incubated on ice for 2 h with Pansorbin (Calbiochem) coated with rabbit anti-mouse immunoglobulin (Sigma) and mouse anti-KIR2D antibody DX27, anti-FLAG antibody M2 (Kodak), or control IgG and then washed in Tris-buffered saline (TBS, 50 mM Tris, 150 mM NaCl, pH 8.0) containing 5 mM CHAPS (Sigma) and protease inhibitors²⁶. Biotinylated peptides corresponding to residues ITETESPY*QELQGQRSDVY*SDLNTQRP in the cytoplasmic domain of DAP12 were synthesized, either unphosphorylated or containing phosphate on both Y residues (provided by C. Turck, UCSF). Control unphosphorylated and Y-phosphorylated CD3ζ peptides¹⁸ were a gift from A. Weiss (UCSF). Biotinylated peptides were incubated with lysates from Jurkat or NK clone A6 cells, precipitated with avidin–agarose, and washed in Tris-buffered saline (50 mM Tris, 150 mM NaCl, pH 7.8) containing 1% Nonidet P-40 and protease inhibitors¹⁸. Immunoprecipitates were analysed by western blot²⁷

using anti-ZAP-70 antibody or rabbit anti-Syk specific antiserum¹⁸ (provided by A. Weiss, UCSF).

Cell activation. Ba/F3 cells expressing either KIR2DS2 alone, DAP12 (FLAG epitope tagged) alone, or the KIR2DS2–DAP12 complex were incubated with the indicated antibodies at 4 °C, washed, and then crosslinked with F(ab')₂ goat anti-mouse immunoglobulin for 3 min at 37 °C. Cells were lysed in TBS containing 1% Nonidet P-40 and protease inhibitors. Total cell lysates or immunoprecipitates of DAP12–FLAG with anti-FLAG antibody M2 were analysed by western blot using horseradish peroxidase-conjugated anti-phosphotyrosine antibody 4G10 (UBI). Jurkat cells stably transfected with the NKAT5 cDNA¹⁶ using a retroviral vector²⁴ were transiently transfected by electroporation with human DAP12 cDNA in the pEF-BOS vector or sham-transfected with a control vector²⁵. After 24 h, transfectants were incubated in microtitre plates precoated (5 µg ml⁻¹) with control immunoglobulin or anti-KIR antibody DX27. After 12 h incubation, transfectants were collected and then stained with FITC-conjugated anti-CD69 or control antibody and analysed by flow cytometry²⁸.

Received 15 September; accepted 29 October 1997.

1. Lanier, L. L. NK cells: from no receptors to too many. *Immunity* **6**, 371–378 (1997).
2. Burshtyn, D. N. *et al.* Recruitment of tyrosine phosphatase HCP by the killer cell inhibitory receptor. *Immunity* **4**, 77–85 (1996).
3. Olcese, L. *et al.* Human and mouse killer-cell inhibitory receptors recruit PTP1C and PTP1D protein tyrosine phosphatases. *J. Immunol.* **156**, 4531–4534 (1996).
4. Houchins, J. P., Lanier, L. L., Niemi, E., Phillips, J. H. & Ryan, J. C. Natural killer cell cytolytic activity is inhibited by NKG2-A and activated by NKG2-C. *J. Immunol.* **158**, 3603–3609 (1997).
5. Biassoni, R. *et al.* The human leukocyte antigen (HLA)-C-specific 'activatory' or 'inhibitory' natural killer cell receptors display highly homologous extracellular domains but differ in their transmembrane and intracytoplasmic portions. *J. Exp. Med.* **183**, 645–650 (1996).
6. Mason, L. H. *et al.* The Ly-49D receptor activates murine natural killer cells. *J. Exp. Med.* **184**, 2119–2128 (1996).
7. Olcese, L. *et al.* Human killer cell activatory receptors for MHC class I molecules are included in a multimeric complex expressed by Natural Killer cells. *J. Immunol.* **158**, 5083–5086 (1997).
8. Reth, M. Antigen receptor tail clue. *Nature* **338**, 383–384 (1989).
9. Chan, A. C., Desai, D. M. & Weiss, A. The role of protein tyrosine kinases and protein tyrosine phosphatases in T cell antigen receptor signal transduction. *Annu. Rev. Immunol.* **12**, 555–592 (1994).
10. Baker, E., D'Andrea, A., Phillips, J. H., Sutherland, G. R. & Lanier, L. L. Natural killer cell receptor for HLA-B allotypes, NKBI: map position 19q13.4. *Chrom. Res.* **3**, 511 (1995).
11. Meyard, L. *et al.* LAIR-1, a novel inhibitory receptor expressed on human mononuclear leukocytes. *Immunity* **7**, 282–290 (1997).
12. Wagtmann, N., Rojo, S., Eichler, E., Mohrenweiser, H. & Long, E. O. A new human gene complex encodes the killer cell inhibitory receptors and a related family of monocyte/macrophage receptors. *Curr. Biol.* **7**, 615–618 (1997).
13. Chang, C. *et al.* Molecular characterization of human CD94: a type II membrane glycoprotein related to the C-type lectin superfamily. *Eur. J. Immunol.* **25**, 2433–2437 (1995).
14. Houchins, J. P., Yabe, T., McSherry, C. & Bach, F. H. DNA sequence analysis of NKG2, a family of related cDNA clones encoding type II integral membrane proteins on human natural killer cells. *J. Exp. Med.* **173**, 1017–1020 (1991).
15. Samaridis, J. & Colonna, M. Cloning of novel immunoglobulin superfamily receptors expressed on human myeloid and lymphoid cells: structural evidence for new stimulatory and inhibitory pathways. *Eur. J. Immunol.* **27**, 660–665 (1997).
16. Colonna, M. & Samaridis, J. Cloning of Ig-superfamily members associated with HLA-C and HLA-B recognition by human NK cells. *Science* **268**, 405–408 (1995).
17. Clevers, H., Alarcon, B., Wileman, T. & Terhorst, C. The T cell receptor/CD3 complex: a dynamic protein ensemble. *Annu. Rev. Immunol.* **6**, 629–662 (1988).
18. Iwashima, M., Irving, B. A., van Oers, N. S. C., Chang, A. C. & Weiss, A. Sequential interactions of the TCR with two distinct cytoplasmic tyrosine kinases. *Science* **263**, 1136–1139 (1994).
19. Chan, A. C. *et al.* Differential expression of ZAP-70 and Syk protein tyrosine kinases, and the role of this family of protein tyrosine kinases in TcR signaling. *J. Immunol.* **152**, 4758–4766 (1994).
20. Bléry, M. *et al.* Reconstituted killer cell inhibitory receptors for major histocompatibility complex class I molecules control mast cell activation induced via immunoreceptor tyrosine-based activation motifs. *J. Biol. Chem.* **272**, 8989–8996 (1997).
21. Hayami, K. *et al.* Molecular cloning of a novel murine cell-surface glycoprotein homologous to killer cell inhibitory receptors. *J. Biol. Chem.* **272**, 7320–7326 (1997).
22. Kubagawa, H., Burrows, P. D. & Cooper, M. D. A novel pair of immunoglobulin-like receptors expressed by B cells and myeloid cells. *Proc. Natl Acad. Sci. USA* **94**, 5261–5266 (1997).
23. Takase, K. *et al.* A new 12-kilodalton dimer associated with pre-TCR complex and clonotype-independent CD3 complex on immature thymocytes. *J. Immunol.* **159**, 741–747 (1997).
24. Onihisi, M. *et al.* Applications of retrovirus-mediated expression cloning. *Exp. Hematol.* **24**, 324–329 (1996).
25. Wu, J., Katzas, S. & Weiss, A. A functional T-cell receptor signaling pathway is required for p95vav activity. *Mol. Cell. Biol.* **15**, 4337–4346 (1995).
26. Lanier, L. L., Yu, G. & Phillips, J. H. Co-association of CD3 ζ with a receptor (CD16) for IgG Fc on human natural killer cells. *Nature* **342**, 803–805 (1989).
27. Phillips, J. H. *et al.* CD94 and a novel associated protein (94AP) form a NK cell receptor involved in the recognition of HLA-A, -B, and -C allotypes. *Immunity* **5**, 163–172 (1996).
28. Lanier, L. L. & Recktenwald, D. J. Multicolor immunofluorescence and flow cytometry. *Methods: A Companion to Methods in Enzymology* **2**, 192–199 (1991).

Acknowledgements. We thank F. Bazan and J. Bolen for helpful discussion, C. Saraiya for expert technical assistance, C. Turck and A. Weiss for phosphopeptides and antibodies, J. Ford, T. McClanahan and K. Bacon for cDNA libraries, J. A. Katheiser, G. Burger and M. Andonian for graphics, J. Cupp, E. Callas, E. Murphy and D. Polakoff for flow cytometry, D. Gorman for DNA sequencing and D. Liggett for oligonucleotides. DNAX Research Institute is supported by Schering Plough Corporation.

Correspondence and requests for materials to: L.L.L. (e-mail: lanier@dnax.org). Sequences are deposited in GenBank under accession numbers AF019562, AF019563, and AF024637.

Involvement of p85 in p53-dependent apoptotic response to oxidative stress

Yuxin Yin^{*†‡}, Yasuo Terauchi[§], Gregory G. Solomon^{*}, Shinichi Aizawa^{||}, P. N. Rangarajan[¶], Yoshio Yazaki[§], Takashi Kadowaki[§] & J. Carl Barrett^{*†}

^{*} Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA

[†] Curriculum in Genetics and Molecular Biology, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27599, USA

[§] Third Department of Internal Medicine, University of Tokyo, Tokyo 113, Japan

^{||} Laboratory of Morphogenesis, Kumamoto University Medical School, Kumamoto 860, Japan

[¶] Department of Biochemistry, Indian Institute of Science, Bangalore 560012, India

Reactive oxygen species have damaging effects on cellular components and so trigger defensive responses by the cell^{1,2} and even programmed cell death^{3,4}, although the mechanisms by which mammalian cells transmit signals in response to oxidative damage are unknown. We report here that the protein p85, a regulator of the signalling protein phosphatidyl-3-OH kinase (PI(3)K), participates in the cell death process that is induced in response to oxidative stress and that this role of p85 in apoptosis does not involve PI(3)K. We show that disruption of p85 by homologous recombination impairs the cellular apoptotic response to oxidative stress. Because the protein p53 is required for cell death induced by oxidative damage, we examined the relation between p85 and p53. Using a chimaeric p53 fusion protein with the oestrogen receptor (p53ER) to supply p53 (p53 is induced upon binding of p53ER to oestradiol) in a p53-deficient cell line, we found that p85 is upregulated by p53 and that its involvement in p53-mediated apoptosis is independent of PI(3)K. We propose that p85 acts as a signal transducer in the cellular response to oxidative stress, mediating cell death regulated by p53.

We reasoned that cellular responses to oxidative stress might be mediated by signal-transducing proteins. Hydrogen peroxide is required to mediate signal transduction by the receptor for platelet-derived growth factor⁵, which associates with p85 protein⁶. p85 is an SH2/SH3-domain-containing protein originally identified as a regulator of phosphatidylinositol-3-OH kinase⁷ and possibly of other signal transduction proteins⁸. To study the role of p85 in the cellular response to oxidative stress, we derived primary mouse embryo fibroblasts (MEFs) from mice with a targeted disruption of the p85 alpha gene locus (Y.T. *et al.*, manuscript in preparation). p85^{+/-} MEFs from wild-type mice with an otherwise identical genetic background were used for comparison. As expected, expression of p85 was normal in p85^{+/-} MEFs but was undetectable in p85^{-/-} MEFs (Fig. 1a). The amount of p85 in normal MEFs increased following treatment with 88 µM H₂O₂ (Fig. 1b; compare lanes 3 and 4). However, the concentration of p85 was unaffected by H₂O₂ in p53-deficient MEFs (Fig. 1b; compare lanes 5 and 6), suggesting that there might be a link between p53 function and the response of p85 to oxidative stress. To determine the effect of disrupting p85 on the cell death pathway, p85^{+/-} and p85^{-/-} MEF cells were exposed to H₂O₂. As shown in Fig. 1c, p85^{+/-} cells were susceptible to H₂O₂ treatment, with more than 60% of p85^{+/-} cells being killed after 24 hours' exposure to 88 µM H₂O₂. In contrast, p85^{-/-} MEF cells were resistant to killing by the same dose of H₂O₂.

[‡] Present address: Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA.

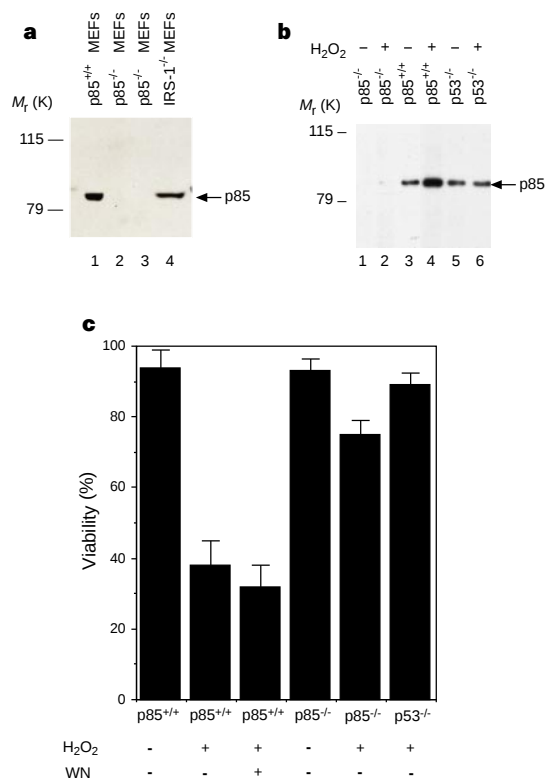


Figure 1 Involvement of p85 in the cellular response to oxidative stress in a manner independent of PI(3)K. **a**, p85^{+/+} MEF and p85^{-/-} MEF cells were isolated and cultured as described in Methods. **b**, MEF cells were exponentially grown and treated with 88 μ M H₂O₂ for 16 h. For **a**, **b**, cells were collected for p85 analysis by immunoblotting using a monoclonal antibody against p85 (1:1,000 dilution) and developed with ECL. **c**, Viability of MEFs after H₂O₂ treatment. Equal numbers (1.2×10^6) of the indicated MEF cells grown in 100-mm dishes in duplicate were exposed to 88 μ M H₂O₂ for 24 h. The presence of 500 nM wortmannin (WN) is indicated. Cell viability was determined by trypan-blue exclusion. Results represent the mean of three independent experiments.

and more than 70% of these cells survived after 24 h of H₂O₂ treatment; about 85% of p53^{-/-} MEFs survived under the same conditions.

To determine the necessity for PI(3)K activity in the cell's oxidative-stress response, we tested the effect of wortmannin, a well characterized PI(3)K inhibitor⁹, on cell viability under oxidative stress. As shown in Fig. 1c, normal MEF cells still underwent oxidative cell death in the presence of wortmannin at 500 nM, a concentration that inhibits PI(3)K activity¹⁰. Wortmannin alone did not cause cell death in p85^{+/+} or in p85^{-/-} MEFs (data not shown). Furthermore, the p85-associated lipid-kinase activity remained unchanged in cells that had undergone cell death as a result of oxidative stress (see below). The morphology of p85^{-/-} cells appeared to be different from the p85^{+/+} cells (Fig. 2a), with p85^{-/-} MEFs being larger than p85^{+/+} MEFs. Most p85^{-/-} cells survived longer treatment with H₂O₂ (Fig. 2b). These results indicate that p85 is involved in the cellular apoptotic response to oxidative damage and that the function of p85 in the cell death pathway does not involve PI(3)K.

Oxidative-stress-induced cell death is strongly influenced by the activity of p53 in the cells. As shown in Fig. 1c, normal MEFs are killed under oxidative stress generated by H₂O₂, whereas MEFs lacking p53 were highly resistant to H₂O₂. To evaluate the role of p53 in the cellular response to oxidative stress, we used a p53ER fusion protein as an inducible system. The complementary DNA encoding human wild-type p53 was fused in-frame with the ligand-binding

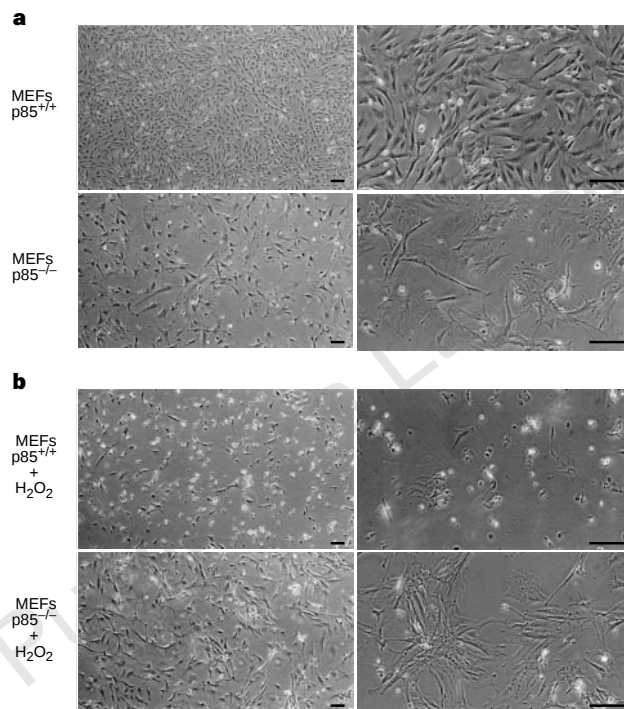


Figure 2 Morphology of p85 MEFs in various conditions. **a**, p85^{+/+} and p85^{-/-} MEFs cells were seeded at a density of 1.2×10^6 in 100-mm dishes and cultured for 24 h. **b**, Cell killing by H₂O₂ in p85^{+/+} MEFs but not in p85^{-/-} MEFs. Exponentially growing MEF cells (1.2×10^6 per 100-mm dish) were exposed to 88 μ M H₂O₂ for 48 h and their morphology photographed. Left panels, magnification $\times 40$; right panels, $\times 100$. Scale bars, 100 μ m.

domain of the human oestrogen receptor in a pCMV mammalian expression vector¹¹. The chimaeric p53ER protein is inactive in the absence of oestradiol but has wild-type p53 function upon binding to oestradiol. This plasmid was then introduced into H1299, a p53-deficient human non-small-cell lung-carcinoma cell line, and stable neomycin-resistant clones expressing p53ER protein were isolated and designated as Hp53ER cells. The amount of p53ER fusion protein, detected by immunoprecipitation (Fig. 3a) as a protein migrating at a position corresponding to a relative molecular mass of $\sim 97,000$ (M_r 97K), was not changed by oestrogen treatment (Fig. 3a, lanes 3–8). To show that the transactivity of wild-type p53 was inducible with the p53ER fusion protein, we used a chloramphenicol acetyltransferase (CAT) reporter construct (PG₁₃CAT) containing a p53 binding consensus sequence upstream of the CAT gene¹². The reporter plasmid was transiently transfected into Hp53ER cells, and CAT activity was evaluated in the presence and absence of 17 β -oestradiol. Figure 3b shows that p53ER strongly induced CAT activity (~ 50 -fold) in the presence of 17 β -oestradiol. Using this p53-inducible system, we examined p85 in an oestrogen-dependent manner. As shown in Fig. 4a, p85 expression was greatly increased in Hp53ER cells in the presence of 17 β -oestradiol alone, and further in the presence of 17 β -oestradiol plus H₂O₂ (Fig. 4a; compare lanes 2 and 3 with 4 and 6). The levels of p85 transcripts were increased following p53 transactivation in Hp53ER cells (data not shown). These results indicate that p85 is regulated by p53 and might be involved in the p53-mediated cellular stress response.

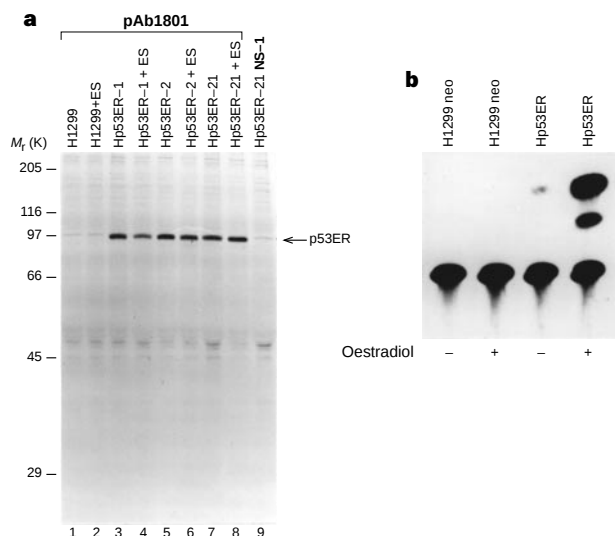


Figure 3 A p53ER inducible system in a p53-deficient human lung-cancer cell line. **a**, Immunoprecipitation analysis of the p53ER fusion protein in the H1299 cell line. Cells were cultured in the absence or presence of 10^{-6} M 17 β -oestradiol (ES) for 16 h. The metabolically labelled proteins were immunoprecipitated by an anti-p53 antibody (pAb 1801, lanes 1–8) or a nonspecific antibody (NS-1, lane 9) and analysed by 10% SDS-PAGE followed by fluorography. **b**, Induction of p53 transactivity by 17 β -oestradiol in Hp53ER cells. The reporter plasmid PG13CAT and β -galactosidase expression vector were co-transfected into 1×10^6 cells using calcium phosphate. Cells were cultured in the absence or presence of 17 β -oestradiol (10^{-6} M) for 16 h and cell lysates prepared for CAT assay.

Because p85 is a regulatory factor of the p110 catalytic subunit of PI(3)K, which phosphorylates inositol-containing lipids, we examined the lipid kinase activity of PI(3)K in our p53-inducible system. As shown in Fig. 4b, the activity of p85-associated lipid kinase was not significantly changed by 17 β -oestradiol and/or H₂O₂ (lanes 1–3), indicating that the p85 function induced in response to oxidative stress may not be due to activation of PI(3)K. We further investigated the significance of p85 in the cellular response to oxidative stress by using a p85 antisense technique in which a cytomegalovirus(CMV)-driven p85 antisense construct was transfected into Hp53ER cells to interfere with the expression of p85. As expected, the levels of p85 varied among the resulting clones (Fig. 5a). PI(3)K activity was decreased in Hp53ER cells expressing antisense p85 (Fig. 4b, lane 4), which contained the lowest level of p85 (Fig. 5a, Hp85AS-7). Oxidative cell death was inhibited in several clones containing low amounts of p85 (Fig. 5a, clones 7 and 9). The induction of cell death by H₂O₂ in Hp85AS-7 and Hp85AS-9 cells was only 11 and 16%, respectively, compared to more than 70% in Hp53ER-Hygro cells (Fig. 5a). Apoptotic cell death mediated by p53, as indicated by DNA ladders, was blocked in Hp53ER/p85AS cells (Fig. 5b: compare lane 4 with lanes 5, 6). These results also indicate that p85 is involved in the p53-dependent apoptotic response to oxidative damage.

We have demonstrated the importance of p85 in the cellular response to oxidative stress. p85 is upregulated by p53 and involved in p53-mediated apoptosis by H₂O₂; p85 regulates the lipid kinase activity of the p110 catalytic subunit of PI(3)K. However, we found that the activity of PI(3)K was unaffected by oxidative stress, indicating that p85 can also function independently of PI(3)K. In addition to binding to p110, p85 also associates with a wide range of membrane-related receptors, receptor tyrosine kinases, cytosolic protein-tyrosine kinases and their substrates^{13–17}. p110 may be involved in signalling for suppression of apoptosis and for cell survival^{18,19}. As a regulator of PI(3)K, p85 may also play a role in these processes in certain conditions or cell types. Because the p85-

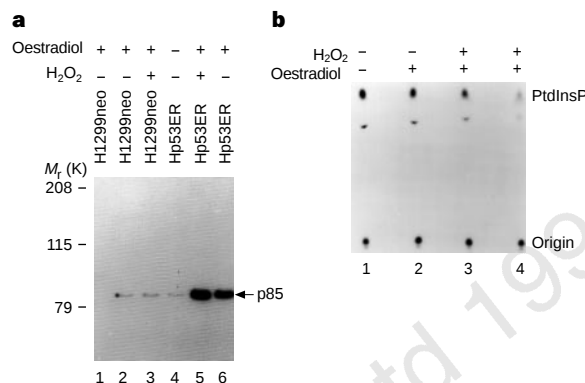


Figure 4 Regulation of p85 in a p53-dependent manner. **a**, Induction of p85 by p53 in an inducible system. H1299 and Hp53ER cells were cultured in the absence or presence of 17 β -oestradiol (10^{-6} M), as indicated, for 16 h. An equal amount (200 μ g) of total protein was precipitated with primary antibody (lane 1, NS-1; lanes 2–6, anti-p85), resolved by 7.5% SDS-PAGE, and transferred onto blotting membranes. Immunoblots were incubated with anti-p85 antibody followed by a peroxidase-conjugated secondary antibody detected with enhanced chemiluminescence. **b**, Thin-layer chromatogram showing p85-associated lipid-kinase activity. Hp53ER cells (lanes 1–3) and Hp53ER/p85AS cells (Hp85AS-7, lane 4) were treated with 10^{-6} M 17 β -oestradiol and/or 130 μ M H₂O₂ for 16 h, as indicated. Cell lysates were immunoprecipitated using anti-p85 antibody for assay of PI(3)K. PtdInsP, phosphatidylinositol phosphate.

associated lipid kinase activity is not induced in response to oxidative stress, p85 may use different pathways to transmit the damage signals that induce cell death.

Our results show that p53 is required for the cellular apoptotic response to oxidative stress by hydrogen peroxide, supporting the idea that p53 function is critical for cell death. It has been proposed²⁰ that p53 induces genes that regulate the redox state of cells undergoing apoptosis and our results are consistent with such a model. In our p53ER inducible H1299 cells, oxidative stress by H₂O₂ was necessary for p53-mediated apoptosis. Although p85 is a downstream mediator of p53-dependent cell death, it is unlikely to be the only one. The observation that p53^{-/-} MEFs are more resistant to oxidative stress than p85^{-/-} MEFs indicates that other signalling pathways may also be involved in p53-mediated oxidative cell death. The link between the tumour-suppressor protein p53 and the SH2/SH3-containing protein p85 is evidence that p53 functions through general cytoplasmic signal transducers which may be important for cell growth and for regulating cell death under different conditions. □

Note added in proof: In p85alpha deficient MEFs, alternative splicing isoform(s) of the p85alpha gene were expressed partially compensating for PI3-kinase activity as in some other tissues (Y.T. *et al.*, submitted for publication). The involvements of these molecules in p53-dependent apoptotic response to oxidative stress needs to be determined.

Methods

Cell lines, reagents and plasmid construction. Mouse embryo fibroblasts were isolated from 12–14-day embryos derived from p85^{-/-} mice and p85^{+/+} mice. IRS-1-deficient MEFs were from mice lacking insulin-receptor substrate-1 (ref. 21). p53^{-/-} MEFs and H1299 have been described^{22,23}. These cell lines were maintained in Earle's minimal essential medium supplemented with 10% fetal bovine serum. The pCMV-p53ER plasmid was constructed as follows: the stop codon of human p53 cDNA was mutated to a HindIII site by site-directed mutagenesis and ligated with the hormone-binding domain of the human

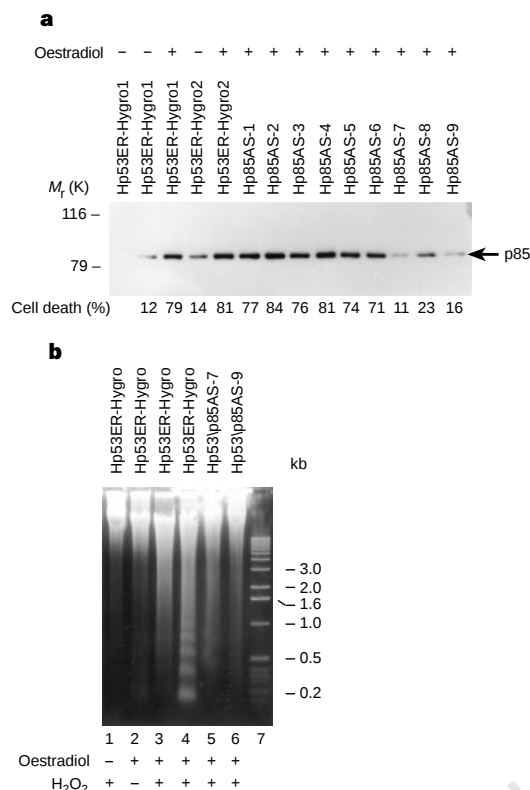


Figure 5 Inhibition of p85 expression and apoptosis in p53ER cells by p85 antisense technology. **a**, Reduction of p85 in Hp53ER cells by p85 antisense expression. Hp53ER cells were transfected with either the pRc/CMV-p85AS construct plus a pMG-Hygro vector for obtaining Hp85AS clones, or pMG-Hygro alone as a control (Hp53ER-Hygro). The growing cells were metabolically labelled with [³⁵S]methionine for 4 h and cell lysates were immunoprecipitated by NS-1 (lane 1) or anti-p85 antibody (lanes 2–14) and separated by 7.5% SDS-PAGE followed by fluorography. For cell killing, exponentially growing cells were treated with H₂O₂ (130 μM) in the absence or presence of 17β-oestradiol (10^{−6} M), as indicated, for 36 h. Cell viability was determined by trypan-blue exclusion. Data represent the average of three independent experiments. **b**, Involvement of p85 for p53-mediated oxidative cell death by apoptosis. Exponentially growing cells of Hp53ER-Hygro (lanes 1–4) and Hp53ER/p85AS (lanes 5, 6) were treated with H₂O₂ (130 μM) in the absence or presence of 17β-oestradiol (10^{−6} M), as indicated, for 6 h (lane 3) or 24 h (other lanes). Genomic DNA was isolated for DNA-ladder analysis.

oestrogen receptor. The p53-ER fusion sequence was then inserted into the BamHI site downstream of the human CMV major intermediate-early promoter region in the expression vector pCMV-Neo-Bam¹¹. To construct a p85 antisense expression vector, full-length human p85 cDNA was ligated in the antisense orientation into a pRc/CMV vector (Invitrogen). The antibodies used were: anti-p53 antibody pAb1801, purchased from Oncogene Science; p85 monoclonal antibody was from UBI; and NS-1, a nonspecific monoclonal antibody, was from ICN.

Gene transfer and CAT assay. Exponentially growing H1299 cells were transfected with constructs using calcium phosphate precipitation and selected with G418 for p53ER clones. Hp53ER cells were transfected with either the pRc/CMV-p85AS construct plus the pMG-Hygro vector to generate Hp85AS, or pMG-Hygro alone as a control (Hp53ER-Hygro). Stable clones were obtained through hygromycin B selection (200 μg ml^{−1}). For CAT assay, PG13CAT plasmid¹² and β-galactosidase expression vector were co-transfected into H1299-p53ER cells. After 24 h of transfection, cells were treated as indicated for 12 h. β-Galactosidase activity was used to standardize the transfection efficiency and CAT was assayed as described²⁴.

Metabolic labelling and immunoprecipitation. Cultures were metabolically labelled with 400 μCi ml^{−1} trans-³⁵S-label (ICN) in methionine-free minimal

essential medium plus 10% dialysed fetal bovine serum for 2 h. Immunoprecipitates were analysed as described²⁵.

Immunoblotting analysis. Exponentially growing cells were lysed in cold Nonidet P-40 in buffer with protease/phosphatase inhibitors. An equal amount of total protein (200 μg per group) was immunoprecipitated with the indicated antibodies. Immunocomplexes were collected on protein A-Sepharose and resolved by 7.5% SDS-PAGE for immunoblotting with a primary antibody against p85, then incubated with peroxidase-conjugated rabbit anti-mouse IgG. Signals were detected with enhanced chemiluminescence (Amersham).

Cellular viability assay and DNA-ladder analysis. Cultured cells were grown at equal density before treatment and collected for counting for viable and non-viable cells by trypan-blue dye exclusion. For electrophoretic analysis of DNA ladders²⁶, growing cells (5 × 10⁵ per 100-mm dish) were treated as indicated. DNA samples (10 μg per well) were electrophoresed in a 1.2% agarose gel and visualized by ethidium bromide staining.

Lipid kinase assay. Equal amounts of total protein (200 μg) from cell lysates from various groups of cells were immunoprecipitated using anti-p85 monoclonal antibody (1:1,000 dilution). Immune complexes were collected with protein A-Sepharose for lipid kinase assay. PI(3)K activity was assayed following incubation with phosphatidylinositol and [γ-³²P]ATP at 30 °C for 1 h. Extracted phospholipids were separated by thin-layer chromatography followed by autoradiography as described²⁷.

Received 21 July; accepted 25 November 1997.

- Panichi, D., Bagchi, M., Hassoun, E. A. & Stohs, S. J. *In vitro* and *in vivo* generation of reactive oxygen species, DNA damage and lactate dehydrogenase leakage by selected pesticides. *Toxicology* **104**, 129–140 (1995).
- Storz, G., Tartaglia, L. A., Farr, S. B. & Ames, B. N. Bacterial defences against oxidative stress. *Trends Genet.* **6**, 363–368 (1990).
- Hockenbery, D. M., Oltvai, Z. N., Yin, X. M., Millman, C. L. & Korsmeyer, S. J. Bcl-2 functions in an antioxidant pathway to prevent apoptosis. *Cell* **75**, 241–251 (1993).
- Korsmeyer, S. J., Yin, X. M., Oltvai, Z. N., Veis Novack, D. J. & Linnet, G. P. Reactive oxygen species and the regulation of cell death by the Bcl-2 gene family. *Biochim. Biophys. Acta* **1271**, 63–66 (1995).
- Sundaresan, M., Yu, Z. X., Ferrans, V. J., Irani, K. & Finkel, T. Requirement for generation of H₂O₂ for platelet-derived growth factor signal transduction. *Science* **270**, 296–299 (1995).
- Hu, P. *et al.* Interaction of phosphatidylinositol 3-kinase-associated p85 with epidermal growth factor and platelet-derived growth factor receptors. *Mol. Cell. Biol.* **12**, 981–990 (1992).
- Carpenter, C. L. & Cantley, L. C. Phosphoinositide kinases. *Curr. Opin. Cell Biol.* **8**, 153–158 (1996).
- Pleiman, C. M., Hertz, W. M. & Cambier, J. C. Activation of phosphatidylinositol-3' kinase by Src-family kinase SH3 binding to the p85 subunit. *Science* **263**, 1609–1612 (1994).
- Arcaro, A. & Wymann, M. P. Wortmannin is a potent phosphatidylinositol 3-kinase inhibitor: the role of phosphatidylinositol 3,4,5-trisphosphate in neutrophil responses. *Biochem. J.* **296**, 297–301 (1993).
- Ferby, I. M., Waga, I., Hoshino, M., Kume, K. & Shimizu, T. Wortmannin inhibits mitogen-activated protein kinase activation by platelet-activating factor through a mechanism independent of p85/p110-type phosphatidylinositol 3-kinase. *J. Biol. Chem.* **271**, 11684–11688 (1996).
- Baker, S. J., Markowitz, S., Fearon, E. R., Willson, J. K. & Vogelstein, B. Suppression of human colorectal carcinoma cell growth by wild-type p53. *Science* **249**, 912–915 (1990).
- Kern, S. E. *et al.* Oncogenic forms of p53 inhibit p53-regulated gene expression. *Science* **256**, 827–830 (1992).
- Otsu, M. *et al.* Characterization of two 85 kd proteins that associate with receptor tyrosine kinases, middle-T/pp60c-src complexes, and PI3-kinase. *Cell* **65**, 91–104 (1991).
- Kotani, K., Hara, K., Kotani, K., Yonezawa, K. & Kasuga, M. Phosphoinositide 3-kinase as an upstream regulator of the small GTP-binding protein Rac in the insulin signaling of membrane ruffling. *Biochem. Biophys. Res. Commun.* **208**, 985–990 (1995).
- Kapeller, R. & Cantley, L. C. Phosphatidylinositol 3-kinase. *Bioessays* **16**, 565–576 (1994).
- Backer, J. M. *et al.* Phosphatidylinositol 3'-kinase is activated by association with IRS-1 during insulin stimulation. *EMBO J.* **11**, 3469–3479 (1992).
- Carpenter, C. L. & Cantley, L. C. Phosphoinositide 3-kinase and the regulation of cell growth. *Biochim. Biophys. Acta* **1288**, 11–16 (1996).
- Yao, R. & Cooper, G. M. Requirement for phosphatidylinositol-3 kinase in the prevention of apoptosis by nerve growth factor. *Science* **267**, 2003–2006 (1995).
- Kauffmann-Zeh, A. *et al.* Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB. *Nature* **385**, 544–548 (1997).
- Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W. & Vogelstein, B. A model for p53-induced apoptosis. *Nature* **389**, 300–305 (1997).
- Tamemoto, H. *et al.* Insulin resistance and growth retardation in mice lacking insulin receptor substrate-1. *Nature* **372**, 182–186 (1994).
- Donchew, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
- El-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, NY, 1989).
- Yin, Y., Tainsky, M. A., Bischoff, F. Z., Strong, L. C. & Wahl, G. M. Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. *Cell* **70**, 937–948 (1992).
- Preston, G. A. *et al.* Induction of apoptosis by c-Fos protein. *Mol. Cell. Biol.* **16**, 211–218 (1996).
- Hartley, D., Meisner, H. & Corvera, S. Specific association of the beta isoform of the p85 subunit of phosphatidylinositol-3 kinase with the proto-oncogene c-cbl. *J. Biol. Chem.* **270**, 18260–18263 (1995).

Acknowledgements. We thank B. Vogelstein for the plasmid PG13CAT; D. Carbone for H1299 cell line; A. Bradley for p53 MEFs; L. T. Williams and E. Y. Skolnik for p85 cDNAs; G. M. Wahl and T. A. Van Dyke for input and advice; G. A. Preston for DNA ladder analysis; J. E. Stenger for critical comments on the manuscript; and S. Sandberg for help with the manuscript. Y. Yin is grateful to T. D. Tlsty and S. T. Lord for their support. This work was supported in part by a Grant-in-Aid for the second term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health and Welfare of Japan (to Y.Y.). Y. Yin was supported by an NIH training fellowship at the University of North Carolina.

Correspondence and requests for materials should be addressed to J.C.B. (e-mail: barrett@nichs.nih.gov).

Location of enhancers is essential for the imprinting of *H19* and *Igf2* genes

Andrea L. Webber, Robert S. Ingram, John M. Levers & Shirley M. Tilghman

Howard Hughes Medical Institute and Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

Genomic imprinting is the process in mammals by which gamete-specific epigenetic modifications establish the differential expression of the two alleles of a gene. The tightly linked *H19* and *Igf2* genes are expressed in tissues of endodermal and mesodermal origin, with *H19* expressed from the maternal chromosome and *Igf2* expressed from the paternal chromosome. A model has been proposed to explain the reciprocal imprinting of these genes¹; in this model, expression of the genes is governed by competition between their promoters for a common set of enhancers. An extra

set of enhancers might be predicted to relieve the competition, thereby eliminating imprinting. Here we tested this prediction by generating mice with a duplication of the endoderm-specific enhancers. The normally silent *Igf2* gene on the maternal chromosome was expressed in liver, consistent with relief from competition. We then generated a maternal chromosome containing a single set of enhancers located equidistant from *Igf2* and *H19*; the direction of the imprint was reversed. Thus, the location of the enhancers determines the outcome of competition in liver, and the strength of the *H19* promoter is not sufficient to silence *Igf2*.

The *H19* and *Igf2* genes lie at one end of a cluster of imprinted genes on mouse chromosome 7 (Fig. 1a). The expression of *H19* and *Igf2* in liver requires the presence of two enhancers located 3' of the *H19* gene^{2,3}. Competition between the genes for these enhancers is biased in favour of *Igf2* on the paternal chromosome by means of germline-inherited DNA methylation, which silences the *H19* promoter⁴⁻⁶. The basis of *H19*'s advantage on the unmethylated maternal chromosome is less clear. From the competition model¹, it can be predicted that repression of expression of the maternal *Igf2* gene should be relieved if the gene is provided with its own enhancers. A similar approach was used to demonstrate that there is competition between the chicken embryonic and adult β -globin

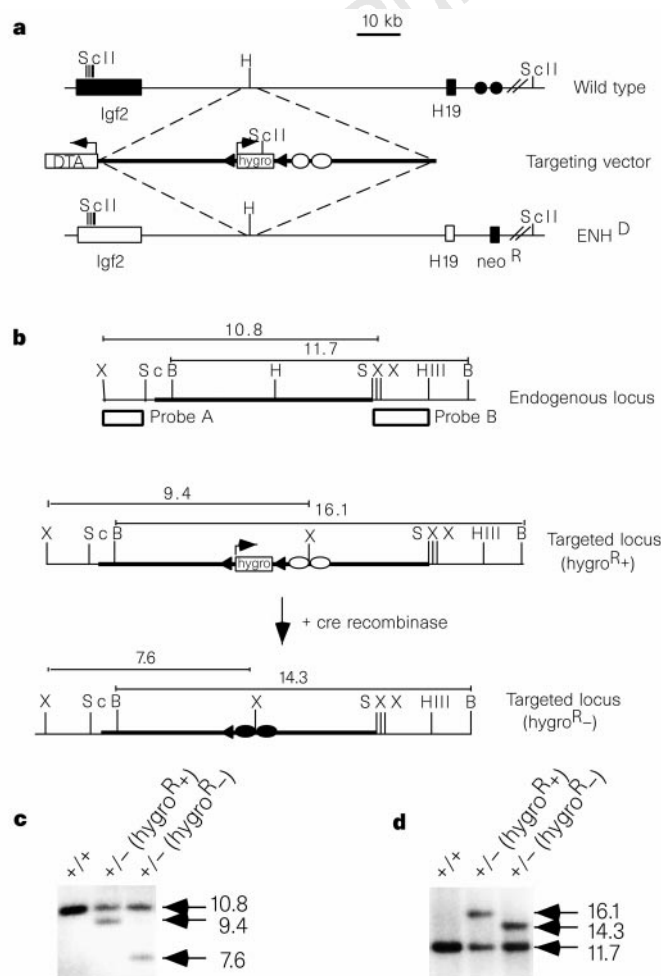


Figure 1 Insertion of the *H19* endoderm-specific enhancers between *Igf2* and *H19*. **a**, Genes are indicated by rectangles, enhancers by circles and ovals, *loxP* sites by closed triangles, and diphtheria toxin gene (DTA) and hygromycin-resistance gene (*hygro*) by labelled open rectangles, with the direction of transcription of the selectable markers indicated by arrows. Restriction enzyme sites are indicated as follows: H, *HpaI*; ScII, *SacII*. **b**, The structures of the endogenous locus and the targeted locus before (*hygro*^R+) and after (*hygro*^R-)

excision of the *hygro*^R gene are shown, with the sizes of the diagnostic *XbaI* and *BstEII* restriction fragments detected with probes A and B. DNA included in the targeting vector is shown as a thick black line. Restriction enzyme sites are as follows: X, *XbaI*; Sc, *SacI*; B, *BstEII*; H, *HpaI*; HIII, *HindIII*. Genomic DNAs were digested with *XbaI* or *BstEII* and hybridized to probe A (**c**) or B (**d**). Wild-type 129Sv/J DNA (+/+) and DNA from a heterozygous targeted cell line (+/-), before and after excision of *hygro*^R, are shown.

genes, which achieve their stage-specific transcriptional regulation through competition for a single intergenic enhancer^{7,8}.

We used gene targeting in embryonic stem (ES) cells to generate mice with a set of endoderm-specific enhancers located roughly equidistant from the *Igf2* and *H19* genes (Fig. 1). A 2.4-kilobase *XbaI*–*BglII* fragment containing the enhancers was integrated 40 kb downstream of *Igf2* and 50 kb upstream of *H19*. The insertion vector included the hygromycin-resistance (*hygro*^R) gene as a selectable marker that was subsequently deleted by transient expression of Cre recombinase⁹. The ES cell line utilized was heterozygous for a deletion of the enhancers (*ENH*^Δ), which were replaced with a neomycin-resistance (*neo*^R) gene³. Using this cell line, we generated two different mutant cell lines, depending on which allele was targeted; a duplication of the *H19* enhancers (*ENH*^{dup}) resulted when the wild-type chromosome was targeted, and a single copy of the *H19* enhancers in a new position (*ENH*^{mov}) resulted when the *ENH*^Δ allele was targeted. We could distinguish between the two outcomes in the targeted ES cells by digesting genomic DNA with *SacII* and hybridizing the DNA fragments to probes derived from the 2.4-kb enhancer fragment and the *neo*^R gene (Fig. 1a).

To examine the effect of duplicating the enhancers, we crossed heterozygous *ENH*^{dup} females with *Mus castaneus* males. Using an allele-specific RNase protection assay, we observed activation of the normally silent maternal allele of *Igf2* in neonatal liver of the progeny (Fig. 2a). No activation was observed in muscle, indicating that the effect in liver was probably dependent on the activity of the endoderm-specific enhancers and not the result of a general disruption in the structure of the locus. In liver, the level of maternal *Igf2* RNA (~30% of paternal *Igf2* RNA) was equivalent to that observed in mice in which the competing *H19* gene and 10 kb of its 5' flank (*H19*^{Δ13}) had been deleted¹⁰. It remains to be determined whether the failure to fully activate *Igf2* in either *H19*^{Δ13} or *ENH*^{dup} liver is a consequence of additional epigenetic modifications of *Igf2* (that must be liver-specific as the gene is fully activated in *H19*^{Δ13} mesoderm) or whether it reflects selection against high levels of hepatic *Igf2*. Nonetheless, the expression of both *H19* and *Igf2* from the maternal allele in *ENH*^{dup} mice concurs with the proposal that the repression of the maternal *Igf2* gene in wild-type animals results from its inability to engage the enhancers.

Maternal expression of *Igf2* was also observed in neonatal gut but not in kidney or lung (Fig. 2a and results not shown). This was unexpected because when a 6.2-kb region that included the enhancers was deleted in *ENH*^Δ mice, a ~70–75% decrease in *Igf2* expression was observed in both kidney and lung³. As the enhancers were defined by transfection into a liver-derived hepatoma cell line², it is possible that the additional 3.8 kb that were deleted in *ENH*^Δ mice contain sequences required for enhancer function in kidney and lung. This theory is consistent with the very low level of expression in kidney¹¹ of transgenes containing the minimal enhancers used here. Alternatively, the intergenic enhancers could be less accessible to the *Igf2* gene in these tissues because of inappropriate DNA methylation. However, the enhancers exhibited patterns of methylation comparable to those of the wild-type enhancers in all tissues examined (data not shown).

The duplication of the enhancers had no effect on the expression of the maternal *H19* gene in liver, as determined by a quantitative RNase protection assay (Fig. 2a). The coexpression of both *H19* and *Igf2* from the same chromosome eliminates the possibility that the production of *H19* RNA in *cis* is sufficient to silence maternal *Igf2* expression. An allele-specific RNase protection assay confirmed that there was no *trans* effect of the duplication and that the normally silent paternal allele of *H19* was repressed in these animals (data not shown).

In the reciprocal cross, in which the mutant chromosome was inherited paternally, no increase in the level of paternal *Igf2* was detected (Fig. 2b). This suggests that *Igf2* expression from the paternal allele cannot be elevated further by the extra set of enhancers. More importantly, the paternal *H19* gene remained silent (Fig. 2b), indicating that silencing of the paternal *H19* gene is governed by DNA methylation and not by competition with *Igf2* for enhancers.

Thus, enhancer competition may be involved in the imprinting of *Igf2*, but it is not known what determines the strong preference for transcription of *H19* from the maternal chromosome. If the inherent strength of the *H19* promoter is much greater than that of the *Igf2* promoter, this could be sufficient to ensure the repression of *Igf2*. Alternatively, the closer proximity of the enhancers to the *H19* gene could achieve the same result. To discriminate between

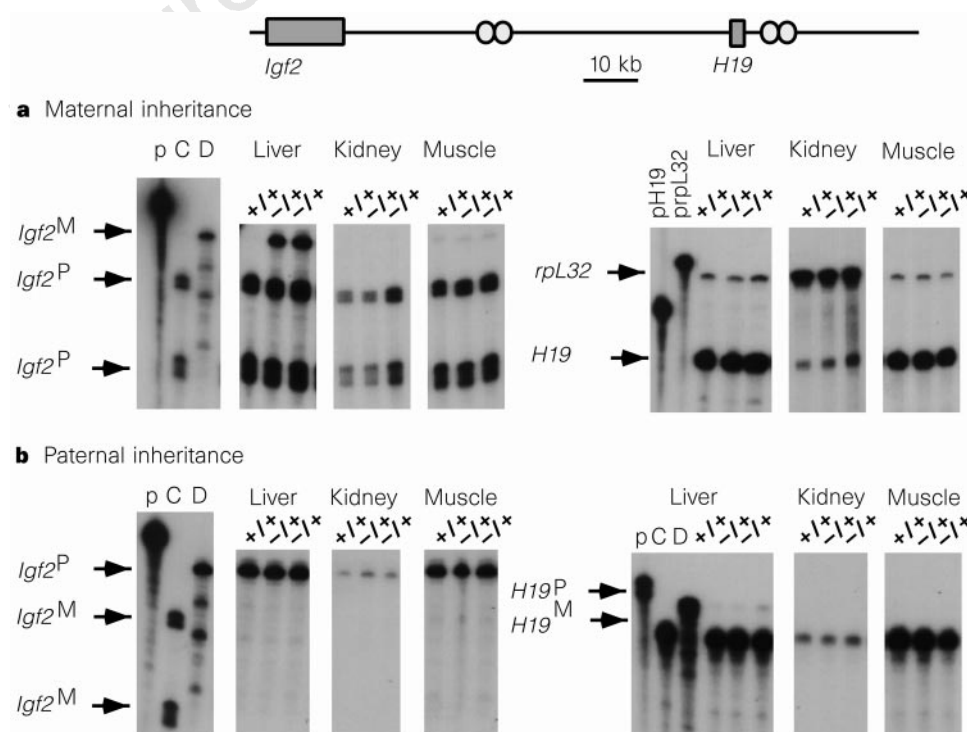


Figure 2 Expression of *Igf2* and *H19* in *ENH*^{dup} mice. The structure of the mutant chromosome is indicated at the top of the figure. Ovals represent enhancers. **a**, Maternal inheritance. Left panel, heterozygous *ENH*^{dup} females were mated with *M. castaneus* males and progeny were killed at 5 days post partum. Liver (5 μg), kidney (15 μg), and muscle (3 μg) RNAs were assayed by allele-specific *Igf2* RNase protection. p, probe; C, *M. castaneus*; D, *M. domesticus* 129SvJ; (+/+), wild-type; (-/+), maternal heterozygote. The migration of the maternal (M) and paternal (P) *Igf2*-protected fragments are indicated on the left. Right panel, the same samples were assayed for *H19* and rpl32 RNAs by non-allele-specific RNase protection. Arrows indicate the rpl32- and *H19*-protected fragments. **b**, Paternal inheritance. Left panel, heterozygous *ENH*^{dup} males were mated with *M. castaneus* females, and the expression of *Igf2* was determined as in **a**. (+/-), paternal heterozygote. Right panel, *H19* expression was assayed using allele-specific RNase protection.

these options, we examined progeny that had inherited the ENH^{mov} allele maternally. As was observed in ENH^{dup} mice, the maternal allele of *Igf2* was expressed in liver and, to a lesser extent, in gut, but not in kidney, lung, or muscle (Fig. 3a, and results not shown). However, the level of *H19* RNA was reduced markedly in liver. In fact, the level of *H19* RNA was nearly identical to the level observed in mice with maternal inheritance of the ENH^A allele³. This suggests that the strength of the *H19* promoter is not the main determinant in the competition for the enhancers. Instead, the preference for *Igf2* transcription in these mice may be determined by the position of the enhancers on the chromosome.

Alternatively, the unexpected absence of *H19* expression in ENH^{mov} maternal heterozygotes could be explained if the mutation induced methylation of the *H19* gene. A *SacI* polymorphism between *Mus domesticus* and *M. castaneus* mice in the 5' flank of the *H19* gene was used to demonstrate that this was not the case; the maternal *M. domesticus*-specific *SacI* fragment was completely digested by the methylation-sensitive restriction enzyme *HpaII* in both wild-type and ENH^{mov} maternal heterozygotes (Fig. 4a). Therefore, the lack of expression of the *H19* gene is not the result of inappropriate DNA methylation.

In mice inheriting the ENH^{mov} allele from their fathers, *Igf2* expression was observed at levels comparable to those seen when ENH^{mov} was inherited on the maternal allele (Fig. 3b). These levels (~30% of wild type) were intermediate between paternal wild-type levels and the low levels seen in ENH^A mice. Thus, in their intergenic position, the enhancers are equally effective on both the maternal and the paternal chromosomes, suggesting that there are no epigenetic differences at the *Igf2* locus that affect the level of expression of the gene.

There are two paternal differentially methylated regions (DMRs) at the *Igf2* gene locus^{12,13}. DMR1 is located ~3 kb upstream of exon 1, and DMR2 is located within the coding region of the gene¹²⁻¹⁴. Although neither region can be a gametic mark, because their allele-specific methylation occurs after fertilization, the paternal methylation of these regions may inhibit the binding of repressors of *Igf2* transcription¹⁵. As *Igf2* is expressed from the ENH^{mov} chromosome whether the chromosome is maternally or paternally inherited, we investigated the methylation status of the DMRs of *Igf2* on the

ENH^{mov} chromosomes. The methylation pattern of DMR1 in heterozygotes with the maternal ENH^{mov} chromosome was the same as the pattern in wild-type mice (Fig. 4b), indicating that the activation of the maternal allele in liver is not accompanied by an increase in its methylation. We also observed the same wild-type pattern of methylation at DMR1 in livers of ENH^{mov} paternal heterozygotes, despite the reduced *Igf2* expression in these animals (Fig. 4b). Similar results were obtained with the DMR2 region (data not shown). Thus, the maternal allele of *Igf2* can be expressed in the absence of methylation at either DMR1 or DMR2.

The preferential expression of *Igf2* in ENH^{mov} maternal heterozygotes has important implications for the role of the *H19* promoter in silencing *Igf2*. If an inherently stronger *H19* promoter is sufficient to set up the repression of *Igf2* on the maternal allele, then *H19* should have been expressed exclusively, regardless of the position of the enhancers. The results rule this possibility out. The strong preference for *Igf2* expression in ENH^{mov} mice could in fact indicate that when the distance of the genes from the enhancers is equalized, *Igf2* has the stronger promoter. However, the enhancers in ENH^{mov} mice are 10 kb closer to *Igf2* than to *H19*, and, therefore, a strict effect of distance could explain the bias in transcription. In either case, our findings indicate that the position of the enhancers relative to the two genes is critical.

Alternatively, *H19* on the maternal allele in ENH^{mov} maternal heterozygotes may be repressed because of a chromatin insulator or boundary element that is normally found between *Igf2* and the enhancers, but, on the ENH^{mov} chromosome, is now located between *H19* and the enhancers. The insulator would act on the wild-type maternal chromosome to insulate *Igf2* from the enhancers. Chromatin boundaries were first described in *Drosophila* as *cis*-acting elements that are capable of insulating a gene and its regulatory elements from position-effect variegation and blocking enhancer-promoter interactions when placed between them^{16,17}.

If there is such an element at the *Igf2/H19* locus, its location must be delineated by *H19*^{Δ13} mutation; that is, it must lie within the region deleted in the mutation, as *Igf2* is transcribed in this mutant. A smaller deletion that removes only the transcription unit and its proximal promoter (*H19*^{Δ3}) results in less pronounced activation of the maternal *Igf2* gene¹⁸, indicating that the 5' flank is a potent

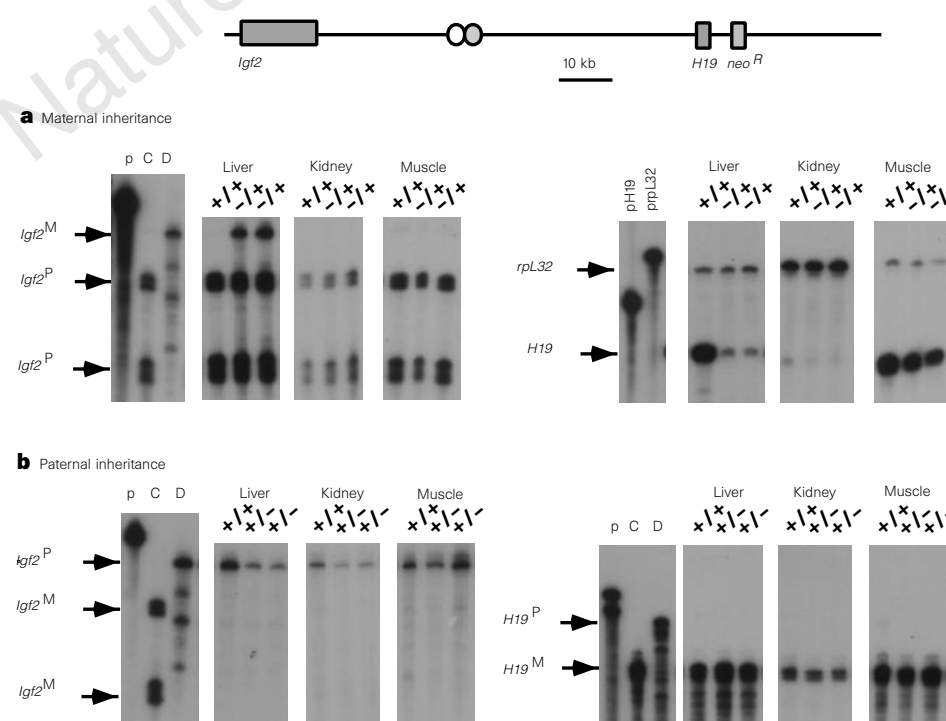


Figure 3 Expression of *Igf2* and *H19* in ENH^{mov} mice. The structure of the mutant chromosome is indicated at the top of the figure. Ovals represent enhancers. **a**, Maternal inheritance. Expression of *Igf2*, *H19*, and *rpl32* RNAs was assayed as in Fig. 2. **b**, Paternal inheritance. Expression of *Igf2* and *H19* RNAs was assayed as in Fig. 2.

determinant in the imprinting of *Igf2*. Interestingly, this same region functions as a silencer in *Drosophila*¹⁹, suggesting that it has unique chromatin properties that can function even across species. A model that includes an insulator on the maternal chromosome requires a mechanism to disable it on the paternal chromosome, thereby

permitting *Igf2* expression. This, presumably, would be the function of the paternal-specific methylation of the 5' flank of *H19*.

Therefore, the proper imprinting of *Igf2* and *H19* on the maternal chromosome depends on the position of the enhancers that the two genes share, at least in liver where the enhancers have been identified. Whether strict promoter competition depends on distance, or whether other chromatin elements affect the outcome, remains to be determined.

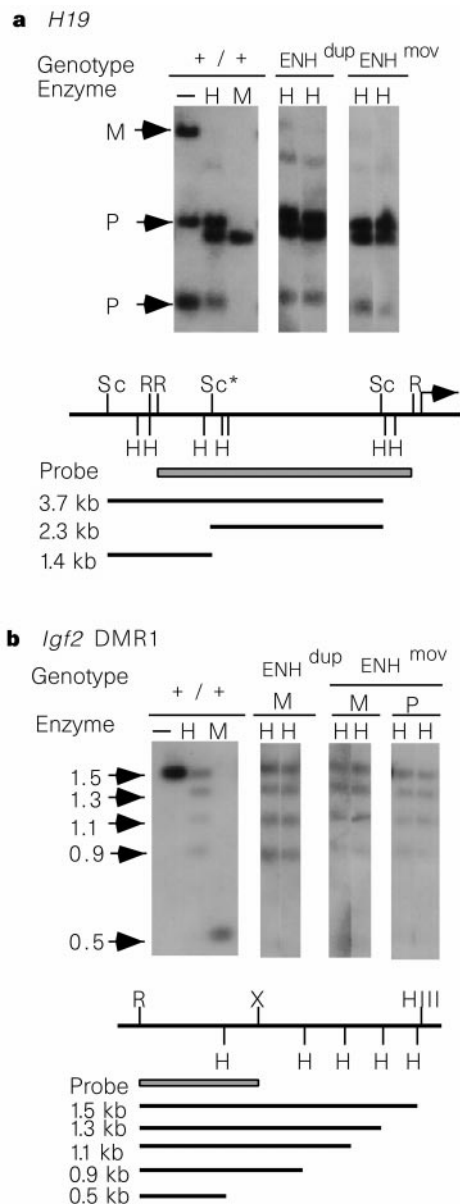


Figure 4 Methylation of *H19* and *Igf2*. **a**, Heterozygous ENH^{dup} or ENH^{mov} females were mated with *M. castaneus* males and progeny were killed at 5 days post partum. Top, liver DNA (15 µg) from the progeny of wild-type animals (+/+) was digested with *SacI* alone (–) or with *SacI* plus *HpaII* (H) or with *SacI* plus *MspI* (M). Arrows at the left of the gels indicate the 3.7-kb *SacI* fragment specific to the maternal (M) allele and the 2.3-kb and 1.4-kb fragments specific to the paternal (P) *M. castaneus* allele. The diagram below the gel illustrates the position of the 3.7-kb *SacI* fragment immediately upstream of the *H19* transcriptional start site (arrow). The hatched box represents the probe; horizontal bars beneath the figure represent digestion products. Restriction enzyme sites are as follows: R, *EcoRI*; H, *HpaII*; all other sites as in Fig. 1. Asterisk denotes polymorphic *M. castaneus*-specific *SacI* site. **b**, Top, liver DNA (15 µg) from the animals in **a** as well as progeny of heterozygous ENH^{mov} males mated with *M. castaneus* females (M) was digested with *EcoRI* and *HindIII* alone (–) or with *EcoRI* and *HindIII* plus *HpaII* (H) or with *HindIII* plus *MspI* (M). Arrows to the left of the gels indicate the sizes and positions of the digestion products. The diagram below the gels illustrates the structure of the 1.5-kb *EcoRI*–*HindIII* fragment from DMR1, as well as the position of the probe and the digestion products.

Methods

Construction of the targeting vector. A *HpaI* site located roughly 40 kb 3' of *Igf2* and 50 kb 5' of *H19* was chosen as the insertion site for targeting. The targeting vector contained a 2.8-kb diphtheria toxin (DTA) negative-selection cassette (ref. 20), 4.6 kb 129/Sv-derived 5'-flanking DNA, a 1.8-kb PGK-hygro^R cassette flanked by *loxP* sites, the 2.4-kb *XbaI*–*BglII* fragment containing the *H19* enhancers, and 4.2 kb of 3' flanking DNA. The vector was linearized at a unique *NorI* site adjacent to the 5'-flanking DNA.

Targeting ES cells. W9.5 ES cells heterozygous for the deletion of the *H19* endoderm enhancers³ were grown on hygro^R primary mouse embryonic fibroblasts²¹. ES cells (1 × 10⁷–3 × 10⁷) were electroporated with a pulse of 240 V/500 µF (Gene Pulser, Biorad) with 25 µg of linearized vector, and were grown after 24 h in 100 µg ml^{–1} hygromycin. Colony DNA was prepared as described²², digested with either *XbaI* or *BstEII* and separated on a 0.8%-agarose gel. DNA fragments were transferred to nitrocellulose filters and hybridized to flanking probes labelled by random hexamer priming²³. Of 167 colonies screened, 47 correctly targeted clones were identified. DNAs from correctly targeted clones were digested with *SacII*, separated by pulsed-field agarose-gel electrophoresis²⁴, transferred to nitrocellulose membranes, and hybridized separately with probes derived from the neo^R gene and the enhancer fragment to ascertain whether the targeting event had occurred on the wild-type or previously targeted chromosome. The PGK-hygro^R cassette was excised by transient expression of Cre recombinase with 25-µg pBS185 Cre plasmid²⁵. Hygro^R was excised from a single targeted clone in which the enhancers were duplicated (9 out of 45 colonies) and a single clone in which the enhancers were moved (4 out of 56 colonies).

Derivation of mice containing intergenic *H19* enhancers. Targeted ES cell clones were injected into C57BL/6 blastocysts and implanted into pseudo-pregnant female mice. To determine whether germline transmission of the mutation had occurred, chimaeras were bred with C57BL/6 mice, and DNA was isolated from tail biopsies of progeny. The DNAs were digested with *XbaI* or *BstEII*; this was followed by Southern blotting and hybridization to radiolabelled probes. Alternatively, the DNAs were analysed by the polymerase chain reaction (PCR) using the forward primer 5'-GAGTAGAGGCTTGATCAGGGC-3' for mice containing hygro^R. Mice with excised hygro^R were identified with the same forward primer and the reverse primer 5'-CCTCACAGCACTACCCTGAGAG-3' from the 5' end of the 2.4-kb enhancer fragment. One line for each clone ± hygro^R was analysed in detail. To detect the ENH^A mutation, DNA was digested with *EcoRI* and hybridized to the 3.0-kb *EcoRI*–*SalI* fragment containing the *H19* structural gene⁹.

RNA isolation and analysis. Total RNA was prepared from neonatal tissues by extraction with LiCl-urea²⁶. RNase protection assays to detect *H19* (refs 11, 27), *rpl32* (ref. 28) and *Igf2* (ref. 10) RNAs were performed as previously described. The levels of maternal *Igf2* RNA were quantified relative to levels of paternal *Igf2* RNA by Phosphorimager densitometry. The levels of *Igf2* in the ENH^{mov} paternal heterozygotes were quantified relative to wild-type levels, using *rpl32* RNA as an internal control.

DNA isolation and methylation analysis. Genomic DNA (15 µg) was digested overnight with *SacI*, *EcoRI* + *HindIII*, *BamHI* (*Igf2* DMR2), or *EcoRI* (*H19* enhancers) in combination with either *HpaII* or *MspI*. The probes were a 4-kb *EcoRI* fragment from the *Igf2* 5' upstream region (for DMR1) (ref. 14), a 904-base-pair *BamHI*–*KpnI* fragment from the coding region of *Igf2* (for DMR2) (ref. 13), and the *XbaI*–*BglII* enhancer fragment.

Received 22 August; accepted 8 December 1997.

1. Bartolomei, M. S. & Tilghman, S. M. Parental imprinting of mouse chromosome 7. *Semin. Dev. Biol.* **3**, 107–117 (1992).
2. Yoo-Warren, H., Pachnis, V., Ingram, R. S. & Tilghman, S. M. Two regulatory domains flank the mouse *H19* gene. *Mol. Cell. Biol.* **8**, 4707–4715 (1988).

3. Leighton, P. A., Saam, J. R., Ingram, R. S., Stewart, C. L. & Tilghman, S. M. An enhancer deletion affects both *H19* and *Igf2* expression. *Genes Dev.* **9**, 2079–2089 (1995).
4. Bartolomei, M. S., Webber, A. L., Brunkow, M. E. & Tilghman, S. M. Epigenetic mechanisms underlying the imprinting of the mouse *H19* gene. *Genes Dev.* **7**, 1663–1673 (1993).
5. Ferguson-Smith, A. C., Sasaki, H., Cattanach, B. M. & Surani, M. A. Parental-origin-specific epigenetic modifications of the mouse *H19* gene. *Nature* **362**, 751–755 (1993).
6. Li, E., Beard, C. & Jaenisch, R. The role of DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
7. Choi, O.-R. B. & Engel, J. D. Developmental regulation of β -globin switching. *Cell* **55**, 17–26 (1988).
8. Nickol, J. M. & Felsenfeld, G. Bidirectional control of the chicken β - and ϵ -globin genes by a shared enhancer. *Proc. Natl Acad. Sci. USA* **85**, 2548–2552 (1988).
9. Gu, H., Zou, Y.-R. & Rajewsky, K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73**, 1155–1164 (1993).
10. Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A. & Tilghman, S. M. Disruption of imprinting caused by deletion of the *H19* gene region in mice. *Nature* **375**, 34–39 (1995).
11. Brunkow, M. E. & Tilghman, S. M. Ectopic expression of the *H19* gene in mice causes prenatal lethality. *Genes Dev.* **5**, 1092–1101 (1991).
12. Sasaki, H. *et al.* Parental imprinting: potentially active chromatin of the repressed maternal allele of the mouse insulin-like growth factor (*Igf2*) gene. *Genes Dev.* **6**, 1843–1856 (1992).
13. Feil, R., Walter, J., Allen, N. D. & Reik, W. Developmental control of allelic methylation in the imprinted mouse *Igf2* and *H19* genes. *Development* **120**, 2933–2943 (1994).
14. Brandeis, M. *et al.* The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *EMBO J.* **12**, 3669–3677 (1993).
15. Walter, J. *et al.* In *Epigenetic Mechanisms of Gene Regulation* (eds Russo, V. E. A., Martienssen, R. A. & Riggs, A. D.) 195–213 (Cold Spring Harbor Laboratory Press, 1996).
16. Kellum, R. & Schedl, P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* **64**, 941–950 (1991).
17. Kellum, R. & Schedl, P. A group of *scs* elements function as domain boundaries in an enhancer-blocking assay. *Mol. Cell Biol.* **12**, 2424–2431 (1992).
18. Ripoché, M.-A., Kress, C., Poirier, F. & Dandolo, L. Deletion of the *H19* transcription unit reveals the existence of a putative imprinting control element. *Genes Dev.* **11**, 1596–1604 (1997).
19. Lyko, F., Brenton, J. D., Surani, M. A. & Paro, R. An imprinting element from the mouse *H19* locus functions as a silencer in *Drosophila*. *Nature Genet.* **16**, 171–173 (1997).
20. McCarrick, J. W. III, Parnes, J. R., Seong, R. H., Solter, D. & Knowles, B. B. Positive-negative selection gene targeting with the diphtheria toxin A-chain gene in mouse embryonic stem cells. *Transgen. Res.* **2**, 183–190 (1993).
21. Johnson, K. A. *et al.* Transgenic mice for the preparation of hygromycin-resistant primary embryonic fibroblast feeder layers for embryonic stem cell selections. *Nucleic Acids Res.* **23**, 1273–1275 (1995).
22. Ramirez-Solis, R. *et al.* Genomic DNA microextraction: a method to screen numerous samples. *Anal. Biochem.* **201**, 331–335 (1992).
23. Feinberg, A. P. & Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**, 266–267 (1984).
24. Chu, G., Vollrath, D. & Davis, R. W. Separation of large DNA molecules by contour-clamped homogenous electric fields. *Science* **234**, 1582–1585 (1986).
25. Sauer, B. & Henderson, N. Targeted insertion of exogenous DNA into the eukaryotic genome by the recombinationase. *New Biol.* **2**, 441–449 (1990).
26. Auffray, C. & Rougeon, F. Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. *Eur. J. Biochem.* **107**, 303–314 (1980).
27. Bartolomei, M. S., Zemel, S. & Tilghman, S. M. Parental imprinting of the mouse *H19* gene. *Nature* **351**, 153–155 (1991).
28. Dudov, K. P. & Perry, R. P. The gene encoding the mouse ribosomal protein L32 contains a uniquely expressed intron containing gene and an unmutated processed gene. *Cell* **37**, 457–468 (1984).

Acknowledgements. We thank T. Watanabe and S. Wang for help in generating chimaeras and the members of our laboratory for discussions. This work was supported by a grant from the National Institute of General Medical Science. S.M.T. is an Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to S.M.T.

Rad23 links DNA repair to the ubiquitin/proteasome pathway

Cherylène Schaubert*, Li Chen*, Prasad Tongaonkar, Irving Vega, David Lambertson, Warren Potts & Kiran Madura

Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, 675 Hoes Lane, Piscataway, New Jersey 08854, USA

* These authors contributed equally to this work.

Rad23 is an evolutionarily conserved protein that is important for nucleotide excision repair^{1–3}. A regulatory role has been proposed for Rad23 because *rad23* mutants are sensitive to ultraviolet light but are still capable of incising damaged DNA^{4,5}. Here we show that Rad23 interacts with the 26S proteasome through an amino-terminal ubiquitin-like domain (Ubl^{R23}). The carboxy terminus of Rad23 binds to the Rad4 DNA repair protein and creates a link between the DNA repair and proteasome pathways. The ultraviolet

sensitivity caused by deletion of the Ubl^{R23} domain may therefore arise from its inability to interact with the proteasome. The fusion proteins glutathione S-transferase (GST)–Rad23 and Rad4–haemagglutinin (HA), and the proteasome subunits Cim3 and Cim5, cofractionate through consecutive chromatography steps. The ubiquitin-like domain of human Rad23 (Ubl^{HRB}) also interacts with the human proteasome. These results demonstrate that ubiquitin-like domains (Ubls) represent a new class of proteasome-interacting motifs.

Rad23 has an unusual N-terminal domain that bears a striking resemblance to ubiquitin¹. This domain, which we call Ubl^{R23}, is important for DNA repair because its elimination causes sensitivity to ultraviolet irradiation¹. Rad23 was isolated in a search for suppressors of the toxicity caused by overexpressing the ubiquitin-dependent N-end rule pathway⁶, suggesting that Rad23 might have a proteolytic function in DNA repair. To explore this idea we linked Rad23 and two truncated mutants to glutathione S-transferase (GST–Rad23, GST– Δ Ubl^{R23}Rad23 and GST–Ubl^{R23}), and immobilized them on glutathione-Sepharose. (GST and GST–Ubl^{R23} were typically expressed at higher levels than GST–Rad23 or GST– Δ Ubl^{R23}Rad23, a mutant lacking the ubiquitin domain). We incubated a western blot containing proteins released from GST and GST–Rad23 beads with antibodies against Cim3 (Sug1) and Cim5 (ref. 7), which are ATPases of the regulatory (19S) subunit of the 26S proteasome⁸. We detected Cim3 (relative molecular mass (M_r) 43K) and Cim5 (54K) in GST–Rad23 beads (Fig. 1a, lanes 2), but not in the control GST beads (Fig. 1a, lane 1). GST–Ubl^{R23} alone could efficiently bind a complex containing Cim3 and Cim5 (Fig. 1a, lane 4), but, a mutant lacking Ubl^{R23} (GST– Δ Ubl^{R23}Rad23) could not (Fig. 1a, lane 3). Two variants of Rad23, bearing small

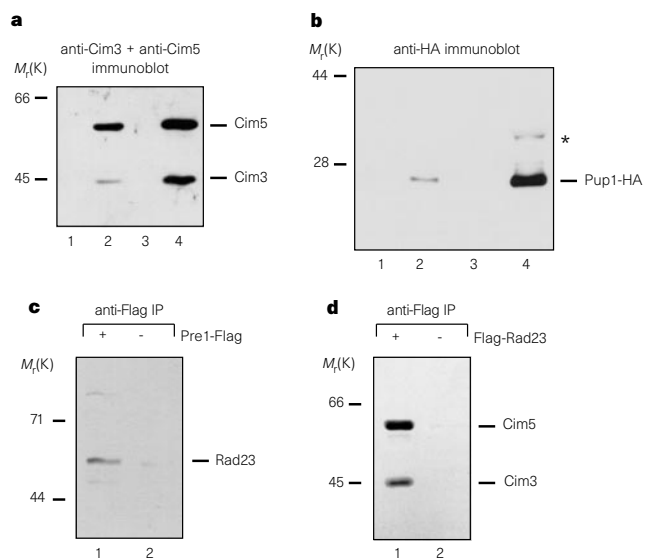


Figure 1 Rad23 interacts with the 26S proteasome. **a**, **b**, GST (lane 1), GST–Rad23 (lane 2), GST– Δ Ubl^{R23}Rad23 (lane 3) and GST–Ubl^{R23} (lane 4) were bound to glutathione-Sepharose. Proteins bound to the beads were resolved on SDS-polyacrylamide gels and transferred to nitrocellulose. **a**, Cim5 (54K) and Cim3 (43K) were detected with specific antibodies in lanes 2 and 4. M_r standards are indicated on the left. **b**, The complex that interacts with GST–Rad23 and GST–Ubl^{R23} contains the 20S subunit Pup1–HA. The additional band (asterisk) may represent a precursor form of Pup1–HA. GST–Ubl^{R23} generally accumulated to much higher levels than GST–Rad23, accounting for the higher levels of Pup1–HA detected in lane 4. **c**, Native Rad23 can be precipitated on Flag–agarose beads in extracts derived from a strain expressing Pre1–Flag (in lane 1), but not from a control extract prepared from a strain lacking Pre1–Flag (lane 2). **d**, Extracts containing Flag–Rad23 can specifically precipitate Cim3 and Cim5 on Flag–agarose beads. These subunits were not recovered from extracts containing a control vector lacking Flag–Rad23 (lane 2).

epitopes on either the N terminus (Flag–Rad23; Fig. 1d) or the C terminus (Rad23–HA; data not shown), also interacted with the proteasome. Both Cim3 and Cim5 were detected in anti-Flag immunoprecipitates prepared from yeast cells expressing Flag–Rad23 (Fig. 1d). Because yeast cells expressing Δ^{Ubl} Rad23 fail to complement *rad23* Δ (ref. 1), these findings suggest that the Rad23–proteasome interaction is important for DNA repair. These data also show that Ubl^{R23} represents a new proteasome interaction signal. A large family of proteins bearing ubiquitin-like extensions has been identified, and our results suggest that they too have proteolytic functions⁹.

To determine whether the GST–Rad23 interacting complex included 20S catalytic subunits we analysed extracts from cells expressing Pre1–Flag (28K) or Pup1–HA (33K), both of which are epitope-tagged derivatives of 20S β -subunits⁸. Both Pup1–HA (Fig. 1b, lane 2) and Pre1–Flag (data not shown) were detected in GST–Rad23 beads after incubation with HA or Flag antibodies, confirming the presence of 20S catalytic subunits. GST–Ubl^{R23} accumulated to higher levels than GST–Rad23, and the recovery of Pup1–HA was proportionately higher (Fig. 1b, compare lanes 2 and 4). To confirm that our findings applied to native Rad23, we immunoprecipitated Pre1–Flag on Flag–agarose beads, resolved interacting proteins on an SDS–polyacrylamide gel, and incubated a western blot with Rad23-specific antibodies. Native Rad23 was readily detectable in immunoprecipitates containing Pre1–Flag but not from a control extract lacking this epitope-tagged proteasome subunit (Fig. 1c). Approximately 5% of cellular Rad23 precipitated with Pre1–Flag (an estimate that is based on the amount of Rad23 that remained on the Flag–agarose beads after 18 h at 4 °C). The *in vivo* interaction could be higher if the interaction with the proteasome is transient or regulated.

The interaction of Rad23 with apparently intact 26S proteasome led us to test whether the complex had ATPase and proteolytic activities. We purified GST–Rad23 interacting proteins on glutathione–Sepharose beads and found significant chymotrypsin activity, consistent with that found in the 20S proteasome (Fig. 2a). We also detected high levels of trypsin and peptidylglutamyl-hydrolytic activities in the GST–Rad23 beads (data not shown), and ATPase activity was detected in the beads containing GST–Rad23 (Fig. 2b). The presence of Cim3 and Cim5 in Rad23 immunoprecipitates strongly suggests that they contribute to this ATPase activity. In agreement with these findings, Flag–Rad23 precipitates also had ATPase and protease activities (data not shown).

Rad23 and Rad4, as well as the human counterparts HHR23-B and XPC, form stable interactions^{2,10}. We therefore tested whether GST–Rad23 interacts with components of the DNA repair and proteolytic pathways. We linked Rad4 to the HA epitope (Rad4–HA) and found that it complemented *rad4* Δ (data not shown). GST–Rad23 and Rad4–HA were expressed simultaneously in yeast cells and were metabolically labelled with [³⁵S]methionine. Radiolabelled extracts were applied to glutathione–Sepharose and bound proteins analysed by SDS–PAGE and fluorography; Rad4–HA was found to interact with GST–Rad23 (Fig. 3, lane 2). The interaction of Rad4–HA with GST–Rad23 did not require Ubl^{R23} (lane 3), demonstrating that distinct regions of Rad23 interact with the proteolytic and DNA repair pathways (compare Figs 1a and 3). Identical samples were transferred to nitrocellulose and analysed by incubation with anti-HA antibodies and, consistent with these findings, Rad4–HA was detected only in lanes 2 and 3 (data not shown). These findings are in agreement with a recent report showing that 21 C-terminal residues in Rad23 are important for interaction with Rad4 (ref. 11). To characterize the interaction of Rad4 with Rad23 we prepared extracts from cells expressing both GST–Rad23 and Rad4–HA and separated proteins on Sephacryl S-200. GST–Rad23 was detected in the void volume coincident with dextran blue, and also in fractions corresponding to its predicted monomeric size (~80K). Rad4–HA, Cim5 and GST–Rad23 could

each be detected in the high-molecular-weight fraction, suggesting that they are components of a single complex (Fig. 4a–c, lane 2). To investigate this further, proteins in the Sephacryl S-200 void volume were chromatographed on Mono-Q. Rad4–HA, Cim5 and GST–Rad23 were detected in fractions eluting at ~0.35 M KCl. These fractions were previously shown to contain catalytically active 26S proteasome¹². Fractions that eluted between 325 and 375 mM KCl from the Mono-Q column were pooled (Fig. 4a–c, lane 3) and chromatographed on Mono-S. Rad4–HA and Cim5 again cofractionated with GST–Rad23 (Fig. 4a–c, lanes 15–18), and a peak of ATPase activity copurified with the GST–Rad23 interacting complex (Fig. 4d).

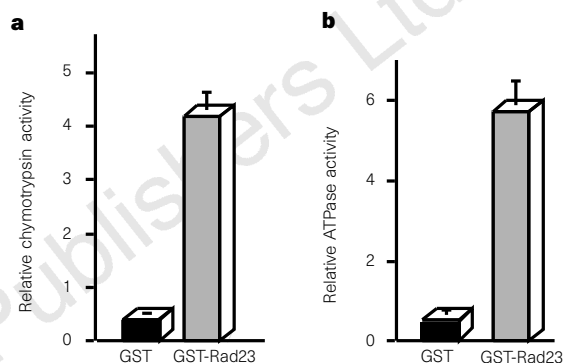


Figure 2 Protease activity associated with Rad23. GST and GST–Rad23 were purified on glutathione–Sepharose and ATPase and protease activities measured. **a**, Chymotrypsin activity was detected in GST–Rad23, but not GST, beads. Data represent the average of three independent measurements. **b**, ATPase activity was detected only in GST–Rad23 containing beads. The data represent four independent measurements (each performed in duplicate), and error bars indicate standard deviation.

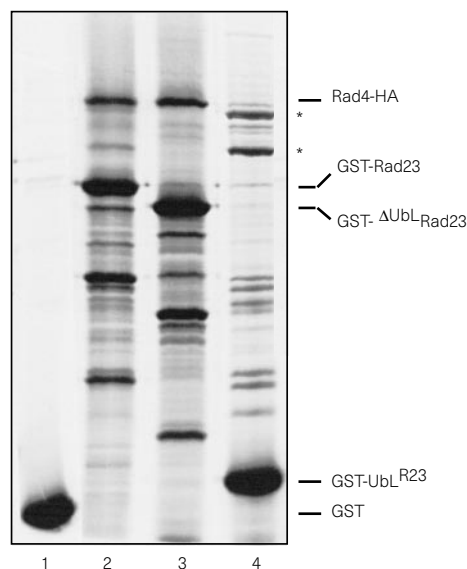


Figure 3 Rad4–HA interacts with Rad23. The positions of [³⁵S]GST fusion proteins and [³⁵S]Rad4–HA are shown. Yeast strains simultaneously expressing Rad4–HA and each of the GST fusion proteins were metabolically labelled with [³⁵S]methionine for 10 min. Extracts were prepared and adsorbed to glutathione–Sepharose. Beads were washed extensively and bound proteins resolved by SDS–PAGE and detected by fluorography. Positions of the GST fusion proteins are indicated (lane order as in Fig. 1a). Rad4–HA is detected in lanes 2 and 3, demonstrating that it interacts with C-terminal sequences in Rad23. Non-specific interactions of other cellular proteins with GST–Ubl^{R23} are indicated by asterisks. Bands of lower *M_r* in lanes 2 and 3 correspond to degradation products of GST–Rad23 and GST–Ubl^{R23}.

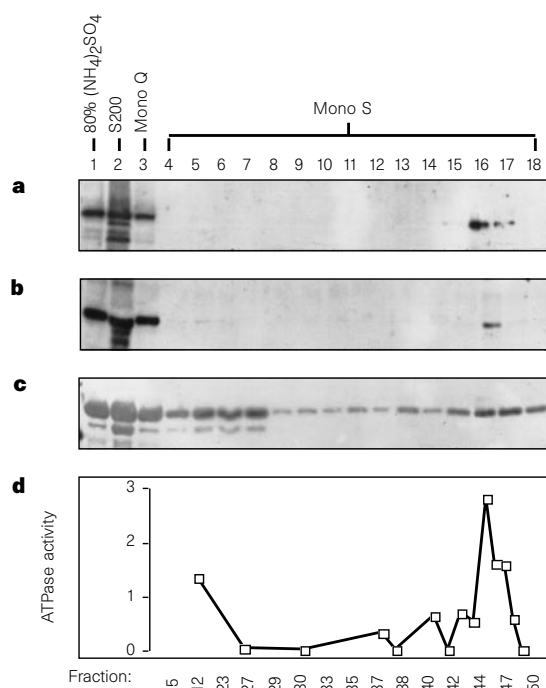


Figure 4 A complex of high M_r contains GST-Rad23, Rad4-HA and Cim5. Yeast extracts were resolved through Sephacryl S-200, Mono-Q and Mono-S. Mono-S fractions were incubated with glutathione-Sepharose and bound proteins separated by SDS-PAGE. A western blot was treated sequentially with antibodies against Rad4-HA (a), Cim5 (b) and Rad23 (c). d, ATPase activity corresponding to Mono-S fractions.

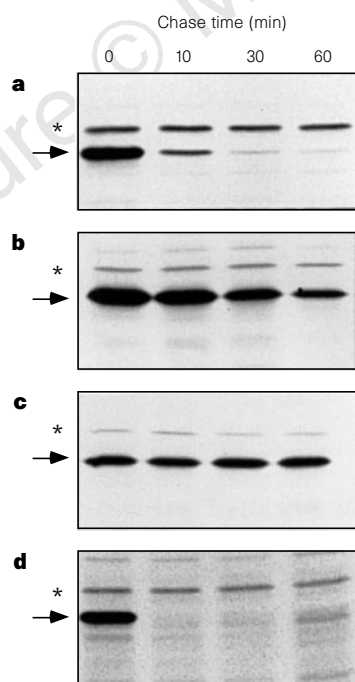


Figure 5 Rad23-HA is degraded by the proteasome. The *in vivo* stability of Rad23-HA in proteasome and vacuolar mutant strains is shown. a, Wild type; b, *pre1-1 pre2-2*; c, *cim5-1*; and d, *pep4Δ prb1Δ*. An arrow indicates the position of Rad23-HA. A ~70K protein that binds Sepharose nonspecifically is indicated by the asterisks.

We found that specific alleles of Rad23 are rapidly degraded, but can complement *rad23Δ*. Minor perturbations of the C terminus of Rad23 (caused by adding a small epitope, or by deleting ~10 residues) resulted in extreme instability ($t_{1/2}$ ~2 min; Fig. 5a). To characterize the degradation of these variants we measured their stability in yeast strains bearing mutations in either proteasome subunits (*pre1-1 pre2-2*, *cim5-1*) or vacuolar proteases (*pep4Δ prb1Δ*). Rad23-HA (an epitope-tagged derivative) was strongly stabilized in *pre1-1 pre2-2* ($t_{1/2}$ ~1 h; Fig. 5b). We also found that Rad23-HA was very stable in *cim5-1* ($t_{1/2}$ > 10 h; Fig. 5c). In contrast, the stability of Rad23-HA was unaffected in *pep4Δ prb1Δ*, which is defective in vacuolar proteolysis¹³ (Fig. 5d). Thus Rad23-HA degradation requires the 26S proteasome. This result raises the possibility that proteasome mutants might be defective in DNA repair. We measured the ultraviolet sensitivity of *cim3-1* and *cim5-1* cells at the permissive temperature and observed a five- to tenfold decrease in survival (data not shown). The finding that both Δ^{Ubl} Rad23 and proteasome mutants contribute to ultraviolet sensitivity strengthens the connection between DNA repair and the ubiquitin-proteasome pathway. A role for protein degradation is also suggested by the rapid degradation of Rad23 variants.

Two human homologues of Rad23, HHR23-A and HHR23-B (ref. 2), contain N-terminal ubiquitin-like domains, suggesting that they act in a similar way to the yeast protein. HHR23-B forms a stable interaction with XPC, the human counterpart of Rad4. We linked the ubiquitin-like domain of HHR23-B (Ubl^{HRB}) to GST to explore the functional relatedness among this class of proteins. GST-Ubl^{HRB} was purified from yeast and incubated with HeLa cell nuclear (Fig. 6) and S100 cytoplasmic extracts (data not shown). We used Cim5-specific antibodies and detected an interaction between GST-Ubl^{HRB} and Mss1 (50K)⁷, the human counterpart of Cim5 (54K) (Fig. 6, lane 2, and data not shown). The interaction between

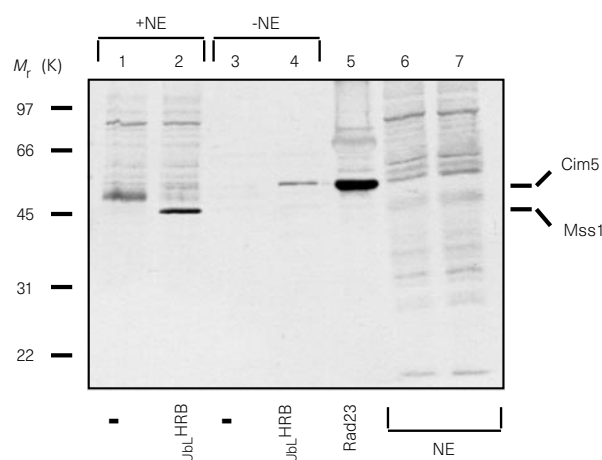


Figure 6 Human HHR23-B interacts with Mss1. GST(-) and GST-Ubl^{HRB} (Ubl^{HRB}) were purified from yeast cells and incubated with HeLa cell nuclear extracts (+/- NE). Mss1 interacted with Ubl^{HRB} (lane 2), and was detected with Cim5 antibodies which were previously shown to cross-react with Mss1 (ref. 7). As a control, the interaction of Cim5 with GST-Rad23 is shown in lane 5. HeLa nuclear extracts that were incubated with GST did not co-precipitate Mss1 (lane 1). GST-Ubl^{HRB} alone shows a small but detectable amount of Cim5 contamination (lane 4), as this fusion protein was purified from yeast. Notice that GST, which was also purified from yeast, does not show this 54K (Cim5) band (lane 3). Under the conditions used, Mss1 was not detected in the nuclear extract without co-precipitation with GST-Ubl^{HRB} (lanes 6 and 7).

GST–Rad23 and Cim5 (lane 5) was used as a control for the antibody reaction.

The biochemical function of Rad23 is poorly understood. Studies using reconstituted protein complexes have shown that Rad23 promotes the assembly of various repair factors^{10,14,15}. Rad23 interacts with TFIH and Rad14 (ref. 5), and can stabilize the preincision complex¹⁶. Furthermore, Rad23 from both yeast and humans (HHR23-B) forms a stable complex with Rad4 (XPC in humans)^{2,5}. It has been shown that purified Rad4 (XPC) is always isolated as a heterodimer containing Rad23 (HHR23-B). These results are consistent with the report that HHR23 regulates the activity of XPC¹⁵. Unfortunately, the use of different experimental protocols to measure DNA repair in *rad23* has yielded highly discordant results. Although the intermediate survival rate of *rad23* is well accepted, some reports have suggested that *rad23* removes dimers very inefficiently¹⁷, whereas others showed a substantial level of repair¹⁵. Use of a reconstituted system showed that a specific DNA adduct was excised in the complete absence of XPC/HHR23-B¹⁴. The primary DNA lesions induced by ultraviolet irradiation are 6-4 photoproducts and cyclobutane pyrimidine dimers^{17,18}. The removal of these lesions, and repair synthesis in minichromosomes¹⁹, have been reported to be inhibited in *rad23*.

A regulatory role for Rad23 is suggested by its known functions, including modulation of Rad4 activity, intermediate ultraviolet sensitivity, its requirement for repair-complex assembly, and our finding that it interacts with the proteasome. Various studies of the role of Rad23 suggest that its activities might be influenced by the type of lesion, and by subtle effects of growth and strain background. In addition to its role in the assembly of repair proteins, Rad23 may also regulate disassembly of the complex. Specifically, the UbL^{R23}-mediated interaction of Rad23 with the proteasome might promote its degradation, causing dissociation of the repair complex. This feature alone might account for several discrepancies between *in vitro* and *in vivo* results. We also propose that the degradation of Rad23 could regulate the recycling of the repair complex, which could, at least in part, explain the repair deficiency in *rad23*. In support of this idea we found that native Rad23 is ubiquitinated and degraded during the transition between the G1 and S phases of the cell cycle (data not shown). □

Methods

Recombinant methods. The genes *RAD23*, *RAD4* and *PUP1* were amplified by polymerase chain reaction (PCR) from yeast strain YPH500 (ref. 6) and cloned into *Eco*R1- and *Kpn*I-digested pKM1362-2 (ref. 6) to generate Rad23–HA, Rad4–HA and Pup1–HA. Pre1–Flag was provided by J. Dohmen. Appropriate oligonucleotides were used to generate UbL^{R23} (codons 1–77 and ΔUbL^{R23}Rad23 (codons 78–398). UbL^{HRB} was generated by PCR from a cDNA template. *RAD23* fragments were linked to GST in pCBGST1 (ref. 20) and expressed by adding 0.1 mM CuSO₄ to the growth medium.

Preparation of protein extracts and immunological methods. To demonstrate co-purification of DNA repair and proteasome subunits, yeast strain LCY156 (MHY501 (ref. 21) expressing GST–Rad23 and Rad4–HA) was grown to late-logarithmic stage in synthetic medium, pelleted and frozen at –70°C. Yeast cells were suspended in lysis buffer (25 mM HEPES, pH 7.5, 100 mM potassium acetate, 1 mM EDTA and 20% glycerol) containing 10 mM β-mercaptoethanol. A mixture of protease inhibitors (including PMSF, leupeptin, aprotinin, antipain and chymostatin) was added immediately before lysis by glass-bead disruption. Protein extracts were precipitated with 80% ammonium sulphate, dissolved in 1 ml gel-filtration buffer (1 mM EDTA, 5 mM MgCl₂, 25 mM Tris–HCl, pH 7.5, 50 mM KCl and 10% glycerol), and separated through Sephacryl S-200. The void volume was collected and chromatographed on FPLC Mono-Q. Fractions corresponding to 325–375 mM in a linear KCl gradient were pooled, diluted and chromatographed on Mono-S. A linear gradient of KCl was applied to Mono-S and fractions collected and applied to glutathione–Sephacryl to allow binding of GST–Rad23 and associated proteins. The samples were boiled for 3 min in sample buffer containing SDS and

resolved on SDS–polyacrylamide gels. Proteins were transferred to nitrocellulose, blocked in 5% milk and incubated with antisera at 1:1,000 (anti-Rad23, anti-HA and anti-Flag) or 1:5,000 dilution (anti-Cim3 and anti-Cim5), and proteins were detected by chemiluminescence (Amersham). To test for interactions between GST and Flag-linked Rad23 derivatives and proteasome subunits, yeast cells were lysed as described above and extracts applied directly to glutathione–Sephacryl (or Flag–agarose). Proteins bound to these beads were separated on two SDS–polyacrylamide gels, one of which was transferred to nitrocellulose, and the other was stained with Coomassie blue to estimate the expression levels of the GST fusion proteins.

ATPase and protease assays. The GST fusion proteins were adsorbed to glutathione–Sephacryl beads, washed and assayed for protease and ATPase activities^{22,23}. Protease substrates included N-CBZ–Leu–Leu–Glu–β-naphthylamide (PGP-like activity), N-CBZ–Ala–Arg–Arg–4-methoxy–β-naphthylamide (trypsin-like activity) and N-CBZ–Gly–Gly–Leu–p-nitroaniline (chymotrypsin-like activity).

Protein stability. Pulse-chase measurements were as described⁶. Briefly, yeast cells expressing Rad23–HA were labelled with [³⁵S]methionine and [³⁵S]cysteine for 5 min. The cells were washed in chase medium containing excess cold methionine and cysteine and 0.5 mg ml^{–1} cycloheximide, and samples were withdrawn at intervals and analysed by immunoprecipitation.

Received 21 July; accepted 24 October 1997.

- Watkins, J. F., Sung, P., Prakash, L. & Prakash, S. The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol. Cell. Biol.* **13**, 7757–7765 (1993).
- Masutani, C. *et al.* Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast *RAD23*. *EMBO J.* **13**, 1831–1843 (1994).
- van der Spek, P. J. *et al.* XPC and human homologs of *RAD23*: intracellular localization and relationship to other nucleotide excision repair complexes. *Nucleic Acids Res.* **24**, 2551–2559 (1996).
- Miller, R. D., Prakash, L. & Prakash, S. Defective excision of pyrimidine dimers and interstrand crosslinks in *rad7* and *rad23* mutants of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **188**, 235–239 (1982).
- Guzder, S. M., Bailly, V., Sung, P., Prakash, L. & Prakash, S. Yeast DNA repair protein *RAD23* promotes complex formation between transcription factor TFIH and DNA damage recognition factor *RAD14*. *J. Biol. Chem.* **270**, 8385–8388 (1995).
- Madura, K. & Varshavsky, A. Degradation of Gα by the N-end rule pathway. *Science* **265**, 1454–1458 (1994).
- Ghislain, M., Udvardy, A. & Mann, C. *Saccharomyces cerevisiae* 26S protease mutants arrest cell division in G2/metaphase. *Nature* **366**, 358–361 (1993).
- Coux, O., Tanaka, K. & Goldberg, A. L. Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* **65**, 801–847 (1996).
- Garrett, K. P. *et al.* Positive regulation of general transcription factor SIII by a tailed ubiquitin homolog. *Proc. Natl Acad. Sci. USA* **92**, 7172–7176 (1995).
- Guzder, S. N., Habraken, Y., Sung, P., Prakash, L. & Prakash, S. Reconstitution of yeast nucleotide excision repair with purified Rad proteins, Replication Protein A, and transcription factor TFIH. *J. Biol. Chem.* **270**, 12973–12976 (1995).
- Wang, Z. *et al.* The *RAD7*, *RAD16* and *RAD23* genes of *Saccharomyces cerevisiae*: Requirements for transcription-independent nucleotide excision repair in vitro and interactions between the gene products. *Mol. Cell. Biol.* **17**, 635–643 (1997).
- Rubin, D. M. *et al.* Identification of the *gal4* suppressor Sug1 as a subunit of the yeast 26S proteasome. *Nature* **370**, 655–657 (1996).
- Hicke, L. & Riezman, H. Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. *Cell* **84**, 277–287 (1996).
- Mu, D., Hsu, D. S. & Sancar, A. Reaction mechanism of human DNA repair excision nuclease. *J. Biol. Chem.* **271**, 8285–8294 (1996).
- Sugasawa, K. *et al.* HHR23B, a human Rad23 homologue, stimulates XPC protein in nucleotide excision repair in vitro. *Mol. Cell. Biol.* **16**, 4852–4861 (1996).
- Reardon, J. T., Mu, T. & Sancar, A. Overproduction, purification, and characterization of the XPC subunit of the human DNA repair excision nuclease. *J. Biol. Chem.* **271**, 19451–19456 (1996).
- McCready, S. Repair of 6-4 photoproducts and cyclobutane pyrimidine dimers in *rad* mutants of *Saccharomyces cerevisiae*. *Mutat. Res.* **315**, 261–273 (1994).
- Verhage, R. A., Seeman, A.-M., Lombaerts, M., van de Putte, P. & Brouwer, J. Analysis of gene- and strand-specific repair in the moderately UV-sensitive *Saccharomyces cerevisiae* *rad23* mutant. *Mutat. Res.* **362**, 155–165 (1996).
- Mueller, J. P. & Smerdon, M. J. Rad23 is required for transcription-coupled repair and efficient overall repair in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 2361–2368 (1996).
- Banihmad, C., Banihmad, A. & O'Malley, B. W. A rapid method combining a functional test of fusion proteins *in vivo* and their purification. *BioTechniques* **16**, 194–196 (1994).
- Chen, P., Johnson, P., Sommer, T., Jentsch, S. & Hochstrasser, M. Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MATα2 repressor. *Cell* **74**, 357–369 (1993).
- Heinemeyer, W., Kleinschmidt, J. A., Saidowsky, J., Escher, C. & Wolf, D. H. Proteinase YSC, the yeast proteasome/multicatalytic-multifunctional proteinase-mutants unravel its function in stress-induced proteolysis and uncover its necessity for cell-survival. *EMBO J.* **10**, 555–562 (1991).
- Merrick, W. C. Assays for eukaryotic protein synthesis. *Methods Enzymol.* **60**, 108–123 (1979).

Acknowledgements. We thank C. Mann for antisera against Cim3/Sug1 and Cim5, M. Ghislain for *cim3-1* and *cim5-1* strains; J. Dohmen for the plasmid expressing Pre1–Flag; J. H. Hoeijmakers for cDNA clones expressing HHR23-A and HHR23-B; and D. Reinberg for HeLa cell extracts. This work was supported by a grant from the NIH (K.M.). I.V. was supported by a pre-doctoral fellowship from the NIH.

Correspondence and requests for materials should be addressed to K.M. (e-mail: maduraki@umdj.edu).

Going software in the neural network

Feeling fuzzy about the logic being use on your data? This selection of software offers packages for database development, data acquisition coupled with statistical analysis and systems for scientific imaging.

ACSL BioMed 3.1

From MGA Software

Simulate clinical trials with computer-assisted trial design software

Working closely with several pharmaceutical companies, MGA has incorporated new functionality into its software for simulating clinical drug trials. Pharmacologists can now simulate proposed clinical trials on specific patient populations, such as pediatrics, smokers and the elderly, and gain insight into these designs before initiating the trials. According to MGA, the program can now predict how variations in trial populations, such as having a significant number of elderly persons, will influence the proposed study. Complex populations can be described and incorporated into any number of trial designs using an intuitive, point-and-click user interface. Once a complex population is defined for one trial, it can be used again or modified for future trial designs.

Reader Enquiry No. 100

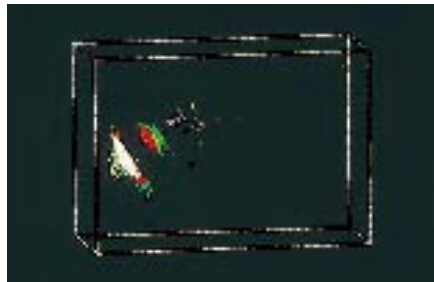
Topkat 5.0

From Oxford Molecular Group

Computational toxicology software developed by Health Designs

This Windows-based program assesses various toxicity metrics based solely on their molecular structures. Because the program can evaluate the specific contribution to a toxicity metric of any user-defined substructural moiety, it can guide a scientist in designing chemical structures having minimal toxicity. By identifying toxic compounds, this technology can also aid in reducing the number of animals used in the toxicity testing of new chemical entities. It features a technology called optimum prediction space (OPS) and integrates a host of algorithms to generate and test the acceptability of predictions and to identify particular molecular substructures for hypothesizing possible molecular sites of toxicity. Topkat is now operating in Windows 3.1 and Windows95 and incorporates an easily-learned interface, extensive on-line help and modules for assessing rodent carcinogenicity, Ames mutagenicity, developmental toxicity potential, rat oral LD₅₀, rat oral chronic LOAEL, skin sensitization (GPMT), fathead minnow LC₅₀, daphnia magna EC₅₀ and LogP. Additional modules are to be released soon for eye and skin irritancy (rabbit), mouse inhalation LC₅₀, rat maximum tolerated dose and aerobic biodegradability.

Reader Enquiry No. 101



C²Diversity samples potential drug compounds.

C²Diversity

From Molecular Simulations

A program that analyses chemical diversity, or similarity, for use in the design and evaluation of compound libraries and reagent sets for combinatorial chemistry

Users can intelligently sample potential drug candidates so that a reduced set can be synthesized and screened. C² Diversity helps to design libraries for synthesis, both at the start of a combinatorial chemistry experiment, and throughout the various iterations of obtaining and using structure activity relationship data to design new libraries. This program helps to reduce the final number of compounds that need to be synthesized, and allows more rational judgement into the combinatorial chemistry process.

Reader Enquiry No. 102

Imaging

Scope-Pro

From Media Cybernetics

Software that automates the movement of microscopes, stages and accessories

This is the latest module for the ImagePro Plus scientific image analysis program. The module is designed to control complex equipment in a simple, repeatable manner. It reduces the need to look through the ocular during image acquisition and to adjust microscope controls manually. The software also automatically controls magnification, focus, motorized filter wheels, shutters, filter sliders and lamps settings, automating repetitive routines for increased efficiency. It supports brands of microscopes and stages that include Zeiss AxioPlan, Olympus Provis, Ludl MAC 2000, Prior H128 series, Sutter Lambda 10-2 filter wheels and shutters and Vincent Associates' shutters. Repetitive functions can be programmed using the Auto-Pro macro or Visual Basic program. By integrating control of the stage into a macro program, users can incorporate any of Image-

Pro's 200 some analysis and processing routines into a fully automated image capture, analysis and data collection process.

Reader Enquiry No. 103

Snapper development kit

From Active Imaging

A software development kit for this range of modular image acquisition boards, which is now available for Macintosh system 7.0

Snapper can acquire images in excess of 25 frames per second, supporting standard or non-standard analog video formats, composite-colour, s-video, RGB and 16-bit digital image acquisition. Triggered acquisition and support for asynchronous reset cameras, such as Pulnix TM-6AS, Hitachi IM-C1, JAI 1550 Progressive Scan, three-chip colour and more, make Snapper well suited for a variety of both industrial and scientific applications. In addition, the digital acquisition module offers specific camera support for an increasing range of cameras, including Pulnix TM9701, Dalsa CL-CX line-scan, Kodak Megaplex 1.4 and Photometrics SenSys. The Mac SDK includes example applications, as well as a Quicktime VDIG, supporting all of the acquisition modules. This allows it to be used immediately with any application supporting Quicktime VDIGs, such as AVID Photoshop, Fusion video recorder, IP Lab Spectrum, and others.

Reader Enquiry No. 104

Image

From Zeiss

A series of image analysis packages that includes a package written specifically for use with Power Macintosh computers

This is a full range of counting, measurement, image, image enhancement and image archiving applications. Users can employ grey-scale or colour acquisition, processing and analysis. This package is said to work seamlessly with other Macintosh-based applications, such as Microsoft Excel and Word, and should be well suited for biological and industrial research applications.

Reader Enquiry No. 105

BioBench

From National Instruments

This is the Virtual Instrumentation Company's software for physiological data acquisition and analysis

This program is offered as a low-cost turnkey package for Windows95/NT and requires no programming to get users up

and running. It can automatically configure and acquire data from many physiological instruments from a variety of manufacturers, as well as acquire data directly from sensors, via the company's plug-in data acquisition hardware. This system interfaces to instruments through a generic BNC connector. Analysis functions include histogram, fast Fourier transform and peak detection. Users can set alarms to trigger when data reaches certain levels. For example, an application monitoring a laboratory subject can trigger an alarm once blood pressure, temperature or respiration reaches a predetermined point.

Reader Enquiry No. 106

Proc-StatXact

From Cytel

A new version of StatXact software that comes as a seamless extension of SAS software

This SAS add-on product allows users of SAS releases 6.11 and 6.12 to perform exact tests on small, sparse or unbalanced data sets with no interruption to processing within the SAS environment. This new capability is said to save time and to ensure greater objectivity and better documentation of analyses. Even more important, the ability to use exact tests automatically rather than chi-square approximations on small or unbalanced safety, efficacy and carcinogenicity tables, should greatly reduce the chance of expensive delays in the review process due to the potential inaccuracy of applying large sample methods to these data sets. Before the release of this package, SAS users who needed to apply the StatXact procedure for computing exact P-values on selected data sets had to delay a SAS production run while they ran the tests using a standalone version. This procedure required users to identify manually suspicious tables from SAS output, enter them into StatXact, run the exact tests and finally insert the results back into their SAS report. This program eliminates these delays and allows users to apply StatXact's high-speed exact and Monte Carlo algorithms to every single table in a large batch run. The Automatic Pilot feature is designed to overcome the risk of hanging up the computer by applying an exact test to a data set that is too large for such processing. It detects data sets that are too large for exact-test processing and automatically switches to Monte Carlo inference.

Reader Enquiry No. 107

Windmill data acquisition

From Windmill

Three new or improved applications that increase the versatility of this data acquisition and control software suite

Applications in this suite include Replay, Alarm Logger and Test-Seq. Replay plays

backlogged data files graphically, allowing the user to scroll through data files looking for significant events. Users can fast forward through uneventful stretches and pause for a more detailed look when something catches their eye. Alarm Logger has been improved with user-specified help messages that can now be shown when an alarm occurs and give instructions on the corrective action to take. Auto-dialing alerts engineers to alarms by telephone, and users can insist on a delay before being alerted to an alarm. This can be useful when monitoring temperatures in a refrigerator where briefly opening and closing a door should not trigger an alarm. There is also password protection to prevent unauthorized users from acknowledging alarms. Test-Seq is a program that interprets an ASCII file of commands and controls measurement and control equipment appropriately. It allows a digital pulse to be output for a specified time period. Other commands let users ramp the values sent to an output channel, monitor and act on alarms, implement control loops, send sequences of keystrokes to other applications, and so on.

Reader Enquiry No. 108

Life Science Workbench

From MDL

This package that comprises Screen and Isis for Excel, providing comprehensive informatics solutions for high-throughput discovery and other aspects of research

Workbench contains flexible and efficient tools for developing, registering and searching methodology information and for capturing, analysing and registering results. The Screen component is designed to accommodate the variety of sample sources available to high-throughput screening laboratories. It manages plate inventory in a hierarchy of storage plates, master plates, daughter plates and assay plates. Users may select any plate format, map samples from each type of plate to the next and transfer the sample to plates with different formats, maintaining a complete audit trail. There is also the ability to track the age and volume of samples in plates, print reports and the uploading of compound, sample or plate information using Screen's flexible file import system.

Reader Enquiry No. 109

Statistica Connectivity kit

From StatSoft

A program that links data acquisition and analysis in a Windows environment

This system is designed to bring statistics and analytical graphics together with external sources of data, such as real-time data acquisition and monitoring systems, measurement devices, data collectors or laboratory equipment. The kit can read data via serial ports or from PC-bus-compatible interface cards,

such as the universal, high-performance Rockwell/Datamyte InterGate card. The kit also includes options to facilitate connecting Statistica to specialized databases that store the relevant information. This integrated system should find use in a variety of environments from simple workstations or small production facilities to large enterprise-wide computing environments. Simple set-up dialogues help with configuration, beginning with how to specify the hardware settings for the source of data and the variables to be used to receive the stream of raw data. Data files that have been configured to receive data from a database or measurement device and can be updated with a single click or keystroke (when data are not imported but received in run time, options are provided to update data continuously or only when requested, to append data or replace them). These updated data files can then be analysed interactively using any part of Statistica or previously created custom scripts.

Reader Enquiry No. 110

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Coming Soon!

SeqWeb™, the new web interface to the popular Wisconsin Package for sequence analysis, is coming in April 1998.

<http://www.gcg.com/seqweb.html>



Genetics Computer Group
608.231.5200 e-mail:info@gcg.com
A wholly owned subsidiary of Oxford Molecular Group, Inc.

READER ENQUIRY NO. 247

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT



PROTEIN SEQUENCING

Membrane-bound, solid or solution samples

£12 per cycle

5 cycles minimum - no set up charge

For further information contact Dr. Joe Gray
Phone: +44 191 222 8612
Fax: +44 191 222 8129
Email: JOE.GRAY@newcastle.ac.uk

READER ENQUIRY NO. 14